Julian Nida-Rümelin

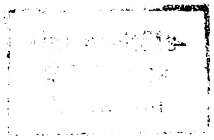
Kritik des Konsequentialismus

416 101 238 500 13

8 93-10069

R. Oldenbourg Verlag München 1993

Gedruckt mit Unterstützung des Förderungs- und Beihilfefonds Wissenschaft der VG Wort.



82200033

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Nida-Rümelin, Julian:
Kritik des Konsequentialismus / Julian Nida-Rümelin. –
München : Oldenbourg, 1993
(Scientia Nova)
ISBN 3-486-56001-8

© 1993 R. Oldenbourg Verlag GmbH, München

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Umschlaggestaltung: Dieter Vollendorf
Gesamtherstellung: R. Oldenbourg Graphische Betriebe GmbH, München

ISBN 3-486-56001-8

-93 P 626

Inhalt

Vorwort 1

Statt einer Einführung. Von Elisabeth Anscombe 3

Teil I: Elemente konsequentialistischer Rationalität 11

1. Kapitel: Konsequenzen 11
§ 1 Handlungskonsequenzen I: Handlungen 12
§ 2 Atom-, Basis- und Abschlußhandlungen 15
§ 3 Savage-Konsequenzen 16
§ 4 Handlungskonsequenzen II: Weltverläufe 20

2. Kapitel: Intentionen 24
§ 5 Intentionalität 24
§ 6 Handlungen als intentionales Verhalten 29
§ 7 Drei Arten von Intentionen 31

3. Kapitel: Bayes'scher Konsequentialismus 36
§ 8 Das Bayes'sche Kriterium 36
§ 9 Konsequentialistische Handlungsgründe 41
§ 10 Propositionale Einstellungen 43

4. Kapitel: Abgrenzungen 46
§ 11 Der eigenorientierte Konsequentialismus 46
§ 12 Der Konsequentialismus als Begründungsmodell 47
§ 13 Intrinsisch wertvolle Handlungen 48
§ 14 Regeln und Konsequenzen 50

Teil II: Konsequentialismus in der Ethik 53

5. Kapitel: Strikter ethischer Konsequentialismus 53
§ 15 Der ethische Konsequentialismus als Spezialisierung konsequentialistischer Rationalität 53
§ 16 Der utilitaristische Idealtypus 54
§ 17 Metaphysische Aspekte 57

6. Kapitel: Gerechtigkeitskonsequentialismus 61
§ 18 Gerechtigkeit im strikten Konsequentialismus 61
§ 19 Deontologischer Konsequentialismus 63

7. Kapitel: Abgeschwächter Konsequentialismus 66
§ 20 Erweiterte Sympathie 67
§ 21 Prärogative 74

8. Kapitel: Folgerungen für die Typologie ethischer Theorien ... 79
§ 22 Methodischer Status 79
§ 23 Subjektiv und objektiv 81
§ 24 Gegenstände 84
§ 25 Teleologisch, deontologisch und konsequentialistisch 86

Teil III: Kritik des Konsequentialismus in der Ethik 89

9. Kapitel: Integrität der Person 89
§ 26 Überbordende Verantwortlichkeit 90
§ 27 Sittlichkeit 92

10. Kapitel: Individuelle Rechte 95
§ 28 Primäre und sekundäre Regeln 95
§ 29 Individuelle Rechte als Beschränkungen 99

11. Kapitel: Das Koordinationsproblem 111
§ 30 Die ideale moralische Person 111
§ 31 Bedingungen der Kooperativität einer Theorie 112
§ 32 Das statische Koordinationsproblem des strikten Konsequentialismus 113
§ 33 Die (statische) Vereinbarkeit von Kooperativität und striktem ethischem Konsequentialismus 115
§ 34 Entkräftung der Gegenbeispiele 117
§ 35 Eine Einschränkung der Vereinbarkeit 124
§ 36 Das dynamische Koordinationsproblem des strikten Konsequentialismus 127

12. Kapitel: Strategische Verfälschung und Manipulation 130
§ 37 Strategische Verfälschung 132
§ 38 Manipulation 136
§ 39 Schlußfolgerungen 137

13. Kapitel: Kritik zweiter Ordnung 139
§ 40 Kritik zweiter Ordnung 139
§ 41 Ethik und Entscheidungsfindung 140
§ 42 Konstruktivismus und Kritik zweiter Ordnung 143
§ 43 Zerstörung normativer Institutionen 144
§ 44 Ziele in intentione recta 147

Teil IV: Kritik konsequentialistischer Rationalität 151

14. Kapitel: Selbstaufhebung konsequentialistischer Rationalität 151

§ 45 Kollektive Irrationalität 152
§ 46 Rationalität und praktische Vernunft 158
§ 47 Strukturelle Irrationalität 165
§ 48 Selbstkontrolle als Orientierung an struktureller Rationalität . 166

15. Kapitel: Handlungsgründe 171
§ 49 Deskriptive und normative Begründungen 171
§ 50 Nicht-konsequentialistische Handlungsgründe 173
§ 51 Kohärenz ohne Konsequentialismus 178
§ 52 Lebensform und Gesellschaftsform 186

Personenregister 189

Sachregister 191

> Es ist aber klar, daß die Ethik nichts mit Strafe und Lohn im gewöhnlichen Sinne zu tun hat. Also muß diese Frage nach den *Folgen* einer Handlung belanglos sein. [...] Es muß zwar eine Art von ethischem Lohn und ethischer Strafe geben, aber diese müssen in der Handlung selbst liegen.
>
> Ludwig Wittgenstein,
> Tractatus logico-philosophicus, 6.422

Vorwort

Vernünftigerweise tut man das, was die besten Folgen hat. Diese scheinbar trivial richtige These ist falsch. Warum sie falsch ist, wird in diesem Buch zu zeigen versucht. Weitgehend offen bleibt, welche Konsequenzen sich daraus ergeben. Ohne Zweifel aber wird die Kritik des Konsequentialismus erst dann wirklich erfolgreich sein, wenn sie nicht in Regressionen auf Kant, Thomas von Aquin, Aristoteles oder eine Fundamentalkritik von Ethik als *theoria*, sondern in eine überzeugende, nicht-konsequentialistische Theorie praktischer Rationalität mündet. Dessen bin ich mir bewußt, und insofern ist dieses Buch unvollständig.

Für wertvolle Anregungen, Kritik und Verbesserungsvorschläge danke ich Wolfgang Stegmüller†, Robert Spaemann, Franz Ferschl, Max Drömmer, Martine Nida-Rümelin, Tania Eden und Dietmar von der Pfordten. Elisabeth Anscombe bin ich dafür dankbar, ihren Text als Einführung verwenden zu dürfen. Mein Dank gilt auch Dagmar Mirbach für die sorgfältige Korrekturarbeit und Sonja Laubach für die Erstellung des Registers sowie dem Förderungs- und Beihilfefonds Wissenschaft der VG Wort für den von ihr gewährten Druckkostenzuschuß.

München, im Dezember 1992 — Julian Nida-Rümelin

Statt einer Einführung

*Von Elisabeth Anscombe**

Bei Moore[1] und den nachfolgenden Ethikern Englands gilt es als ziemlich klar, daß die „richtige Handlung" diejenige Handlung ist, welche die bestmöglichen Konsequenzen nach sich zieht (wobei auch Werte an sich, die von einigen „Objektivisten"[2] bestimmten Arten von Handlungen zugeschrieben werden, zu den Konsequenzen zählen). Hieraus folgt nun, daß ein Mensch, subjektiv gesprochen, gut handelt, wenn er in den gegebenen Umständen nach seinem Urteil über sämtliche Konsequenzen dieser bestimmten Tat im Hinblick auf die besten Konsequenzen handelt. Ich sage, daß dies folge – nicht, daß irgendein Philosoph genau dies gesagt habe. Denn die Diskussion dieser Fragen kann natürlich höchst kompliziert werden: So kann beispielsweise bestritten werden, daß die Ausdrucksweise „dies und dies ist die richtige Handlung" befriedigend sei, mit der Begründung, daß etwas, um ein Prädikat haben zu können, zunächst einmal existieren müsse – wonach die Formulierung „ich bin verpflichtet" vielleicht am besten wäre; ein anderer Philosoph mag bestreiten, daß „richtig" ein „deskriptiver" Ausdruck sei, und dann auf einem Umweg über die Sprachanalyse zu einer Ansicht gelangen, die auf dasselbe hinausläuft wie: „Die richtige Handlung ist diejenige, welche die besten Konsequenzen hervorbringt" (z. B. die Ansicht, man bilde seine „Prinzipien", um das Ziel zu verwirklichen, für dessen Verfolgung man sich entscheidet, wobei der Zusammenhang zwischen „Entscheidung" und dem „Besten" in der Weise gedacht wird, daß eine überlegte Entscheidung bedeutet, seine Handlungsweise so zu wählen, daß man die besten Konsequenzen herbeiführt); ferner muß die Rolle dessen, was man „moralische Prinzipien" nennt, und die Rolle des „Motivs der Pflicht" beschrieben werden; die Unterschiede zwischen „gut", „moralisch gut" und „richtig" sind zu

* Im Jahre 1958 publizierte Elisabeth Anscombe einen Artikel „Modern Moral Philosophy", mit dessen Thesen sie sich in Gegensatz zu fast allen zeitgenössischen Ethikern, insbesondere analytischer Provenienz, stellte. In diesem Artikel wird zum ersten Mal der Terminus Konsequentialismus gebraucht.
Der Beitrag erschien zunächst in *Philosophy* 32 (1958), dann 1969 in *The Is/Ought Question*, hg. von W. D. Hudson, London 1969; ins Deutsche übertragen von F. Scholz in: *Seminar: Sprache und Ethik. Zur Entwicklung der Metaethik*, hg. von G. Grewendorf und G. Meggle, Frankfurt am Main 1974, woraus wir die folgenden Auszüge (mit freundlicher Genehmigung des Suhrkamp Verlages) übernehmen.

[1] Vgl. G. E. Moore, *Principia Ethica*, Cambridge 1903, J.N-R.

[2] Oxforder Objektivisten unterscheiden zwar zwischen „Konsequenzen" und „Werten an sich" und geben sich so den irreführenden Anschein, keine „Konsequentialisten" zu sein. Sie teilen jedoch nicht die Auffassung – und Ross lehnt sie ausdrücklich ab –, daß etwa die Verurteilung eines Unschuldigen etwas so Schwerwiegendes ist, daß es nicht z. B. durch nationale Interessen aufgewogen werden könnte. Ihre Unterscheidung ist somit ohne Belang.

klären, die besonderen Eigenschaften von „sollte"-Sätzen müssen untersucht werden. Derartige Diskussionen erwecken den Anschein, als gebe es zwischen den einzelnen Ansichten signifikante Unterschiede, während doch das eigentlich Signifikante eine grundlegende Gleichartigkeit ist. Die grundlegende Gleichartigkeit wird deutlich, wenn man sich vergegenwärtigt, daß nach der Philosophie der bekanntesten englischen Moralphilosophen die Auffassung unhaltbar ist, nach der es nicht richtig sein kann, Unschuldige – aus welchen Zwecküberlegungen auch immer – zu töten, und daß jemand, der anders denkt, im Irrtum ist. (Ich muß beide Punkte erwähnen; denn während z. B. Hare[3] eine Philosophie lehrt, die einen zu dem Urteil ermutigen würde, daß man sich um übergeordneter Zwecke willen für die Tötung eines Unschuldigen entscheiden „sollte", würde er, wie ich glaube, ebenfalls lehren, daß man demjenigen keinen Irrtum vorwerfen könne, dem es beliebt, es zu seinem „obersten praktischen Prinzip" zu machen, daß ein Unschuldiger um keines Zweckes willen getötet werden darf: das sei eben einfach sein „Prinzip". Mit dieser Ergänzung jedoch, glaube ich, kann man sehen, daß das Gesagte auf jeden einzelnen englischen Moralphilosophen seit Sidgwick zutrifft.) Dies nun ist ein signifikanter Umstand; denn es bedeutet, daß alle diese Philosophien mit der christlich-jüdischen Ethik völlig unvereinbar sind. Denn charakteristisch für diese Ethik war und ist die Lehre, daß gewisse Dinge verboten sind, welche *Konsequenzen* auch immer drohen mögen, so zum Beispiel das Töten eines Unschuldigen um irgendeines auch noch so guten Zweckes willen, stellvertretende Bestrafung, Verrrat (womit ich meine, das Vertrauen eines Menschen in einer wichtigen Sache durch das Versprechen treuer Freundschaft zu gewinnen und ihn dann an seine Feinde zu verraten). Götzendienst, Sodomie, Ehebruch, Ablegen eines falschen Glaubensbekenntnisses. Das Verbot gewisser Dinge allein aufgrund ihrer Beschreibung als so und so identifizierbare Arten von Handlungen, ohne Rücksicht auf weitere Konsequenzen, macht sicher nicht das Ganze der hebräisch-christlichen Ethik aus. Es ist jedoch ein bemerkenswerter Zug dieser Ethik. Und wenn alle Philosophen seit Sidgwick so schreiben, daß diese Ethik ausgeschlossen wird, spräche es für ein gewisses geistiges Provinzlertum, wenn man nicht diese Unvereinbarkeit als das bedeutendste Charakteristikum dieser Philosophen betrachten und im Vergleich dazu ihre Abweichungen untereinander als ziemlich geringfügig ansehen würde.

Bemerkenswert ist, daß keiner dieser Philosophen erkennen läßt, daß er sich der Existenz einer solchen Ethik bewußt ist, zu der er in Widerspruch steht: sie scheinen es alle für offenkundig zu halten, daß ein Verbot wie das des Tötens angesichts gewisser Konsequenzen unwirksam ist. Dabei zielt die Striktheit des Verbots natürlich gerade darauf ab, *sich nicht durch befürchtete oder erhoffte Konsequenzen verführen zu lassen.*

[3] Vgl. R. M. Hare, *The Language of Morals*, London 1952, dt. *Die Sprache der Moral*, Frankfurt/M. 1972, J.N-R.

Wenn man den Übergang von Mill[4] zu Moore betrachtet, wird man vermuten, daß er irgendwann von irgendwem vollzogen wurde; als naheliegend bietet sich Sidgwick an. (...) Vom Standpunkt der gegenwärtigen Untersuchung aus am wichtigsten ist Sidgwicks Definition der Absicht. Er definiert Absicht in einer Weise, daß gesagt werden muß, man beabsichtige jede vorhergesehene Konsequenz einer vorsätzlichen Handlung. Diese Definition ist offensichtlich nicht korrekt, und ich glaube sagen zu können, daß sie heute niemand mehr verteidigen würde[5]. Er verwendet sie, um eine ethische These vorzubringen, die heute weithin Zustimmung fände: die These nämlich, daß es für die Verantwortlichkeit eines Menschen für etwas, was er vorhergesehen hat, keine Rolle spiele, daß er weder als Zweck noch als Mittel zum Zweck gewünscht hat. Wenn wir die Terminologie des Beabsichtigens korrekter verwenden und Sidgwicks falsche Konzeption meiden, können wir die These so formulieren: Für die Verantwortung eines Menschen für eine Wirkung seiner Handlung, die er vorhersehen kann, spielt es keine Rolle, daß er diese Wirkung nicht beabsichtigt. Dies klingt nun ziemlich belehrend; ich glaube, es ist kennzeichnend für eine sehr ungute Degeneration des Denkens über solche Fragen, wenn es belehrend klingt. Wir können sehen, worauf die These hinausläuft, wenn wir ein Beispiel betrachten. Nehmen wir an, ein Mensch sei verantwortlich für den Unterhalt eines Kindes. Er täte also etwas Schlechtes, wenn er dem Kind vorsätzlich die Unterstützung entzöge. Es wäre schlecht, sie ihm deshalb zu entziehen, weil er es nicht weiter unterstützen will; und ebenso wäre es schlecht, sie ihm deshalb zu entziehen, weil er etwa dadurch jemand anderen zu irgendeiner bestimmten Handlung zwingen will. (Wir können für dieses Argument annehmen, daß es für sich selbst genommen durchaus lobenswert wäre, den anderen zu dieser Handlung zu zwingen.) Nun habe er aber die Wahl, entweder etwas Verwerfliches zu tun oder aber ins Gefängnis zu gehen; geht er ins Gefängnis, so folgt, daß er dem Kind seine Unterstützung entzieht. Nach der Lehre Sidgwicks ergibt es keinen Unterschied in der Verantwortung dafür, daß er den Unterhalt des Kindes abbricht, ob er es nun um dieser Tat selbst oder um irgendeines anderen Zweckes willen tut, oder ob es sich als vorhergesehene und unvermeidbare Folge daraus ergibt, daß er ins Gefängnis geht, statt etwas Verwerfliches zu tun. Er muß also die relative Verwerflichkeit des Unterstützungsentzugs einerseits und der verwerflichen Tat anderseits gegeneinander abwägen; und es kann leicht sein, daß die schlechte Tat wirklich weniger unmoralisch wäre, als dem Kind absichtlich die Unterstützung zu entziehen; wenn nun die Tatsache, daß der Unterstützungsentzug ein Nebeneffekt davon ist, daß er ins Gefängnis geht, keine Rolle für seine Verantwortung spielt, wird diese Überlegung ihn

[4] Vgl. J. St. Mill, *Utilitarianism*, in Bd. I der *Collected Works of J. St. Mill*, hg. von F. E. Priestley u. J. M. Robson, Toronto 1963ff.

[5] Vgl. G. E. M. Anscombe, *Intention*, Oxford 1957, dt. *Absicht*, Freiburg/München 1986. J.N.R.

dazu tendieren lassen, die verwerfliche Tat zu begehen, was immer noch etwas ziemlich Schlechtes sein kann. Und wenn er erst einmal begonnen hat, die Sache in diesem Licht zu betrachten, wird es für ihn natürlich das einzig Vernünftige sein, allein über die Folgen nachzudenken und nicht darüber, ob diese oder jene Handlung an sich schlecht ist. So daß er, wenn er mit guten Gründen urteilt, daß kein *großer* Schaden daraus entsteht, etwas viel Verwerflicheres tun kann, als dem Kind vorsätzlich die Unterstützung zu entziehen. Und falls sich seine Überlegungen als falsch erweisen, wird es scheinen, als sei er nicht für die Folgen verantwortlich, da er sie nicht vorausgesehen hat. Denn die These Sidgwicks läuft im Endeffekt darauf hinaus, daß es völlig unmöglich sei, die Schlechtigkeit einer Handlung zu beurteilen, außer im Licht von *erwarteten* Konsequenzen. Aber wenn dem so ist, dann muß jeder die Schlechtigkeit im Licht der von ihm *selbst* erwarteten Konsequenzen beurteilen; und es folgt, daß man sich von den *tatsächlichen* Konsequenzen der verwerflichsten Taten lossprechen kann, solange man geltend machen kann, sie nicht vorhergesehen zu haben. Wohingegen ich behaupten würde, daß ein Mensch für die schlechten Folgen seiner schlechten Handlung verantwortlich ist, die guten hingegen nicht als Verdienst ansehen kann; und daß er umgekehrt keine Verantwortung für die schlechten Folgen seiner guten Handlungen trägt.

Daß Sidgwick – soweit es um Verantwortung geht – jeglichen Unterschied zwischen vorhergesehenen und beabsichtigten Konsequenzen leugnet, hängt bei ihm nicht mit dem Aufbau irgendeiner bestimmten „Methode der Ethik" zusammen; er tat diesen schwerwiegenden Schritt um dieses Schrittes selbst willen und zugunsten aller; und mir scheint es plausibel anzunehmen, daß *dieser* Schritt Sidgwicks den Unterschied erklärt, der zwischen dem herkömmlichen Utilitarismus und jenem von mir so genannten *Konsequentialismus* besteht, der für ihn und für alle englischen Moralphilosophen nach ihm kennzeichnend ist. Durch diesen Schritt erhielt jene Art von Erwägungen, die man früher als Versuchung bezeichnet hätte und die einem von Frauen und schmeichelnden Freunden nahegelegt worden waren, einen festen Status durch die Moralphilosophen und ihre Theorien.

Der Konsequentialismus ist als Philosophie notwendig oberflächlich. Denn in der Ethik gibt es immer Grenzfälle. Als Aristotelianer oder als jemand, der an göttliche Gesetze glaubt, wird man sich angesichts eines Grenzfalles überlegen, ob unter den und den Umständen, so und so zu handeln, z. B. ein Mord oder ein Akt der Ungerechtigkeit wäre; dementsprechend würde man urteilen, ob die Handlung zu tun oder zu unterlassen ist. Das wäre die Methode der Kasuistik; und wenn diese Methode auch dazu verführen kann, in Randfällen einmal ein Auge zuzudrücken, erlaubt sie doch nicht, das Zentrum zu zerstören. Ist man jedoch Konsequentialist, so ist es töricht zu fragen: „Was ist unter den und den Umständen die richtige Handlung?" Der Kasuist stellt eine solche Frage nur,

um zu fragen: „Wäre es *erlaubt*, das und das zu tun?" oder „Wäre es erlaubt, das und das *nicht* zu tun?" Nur wenn es nicht erlaubt wäre, das und das *nicht* zu tun, könnte er sagen: „*Dies muß getan werden*".[6] Andernfalls kann er keine bestimmte Handlung vorschreiben, wenn er auch *gegen* eine Handlung sprechen kann; denn in einem *tatsächlichen* Fall könnten die Umstände (über die bloß vorgestellten Umstände hinaus) auf alle Arten von Möglichkeiten hindeuten, und man kann nicht im vorhinein wissen, welche Möglichkeiten es sein werden. Der Konsequentialist hat nun keine Basis, von der aus er sagen kann: „Dieses wäre erlaubt, jenes nicht"; denn nach seiner eigenen Hypothese sind es die Konsequenzen, die dies entscheiden sollen; und er hat kein Recht, so zu tun, als könne er festlegen, welche möglichen Wendungen ein Mensch dieser oder jener Tat geben kann; bestenfalls kann er sagen, ein Mensch dürfe dieses oder jenes nicht herbeiführen; er hat kein Recht zu sagen, ein Mensch werde in einer konkreten Situation das und das herbeiführen, wenn er nicht so und so handle. Ferner muß der Konsequentialist, um sich überhaupt Grenzfälle vorzustellen, natürlich irgendeine Art von Gesetz oder Standard annehmen, nach denen dies ein Grenzfall ist. Woher aber erhält er seine Standards? In der Praxis lautet die Antwort ständig: von den in seiner Gesellschaft oder in seiner Gruppe geltenden Standards. Und es ist in der Tat kennzeichnend für alle diese Philosophen, daß sie außerordentlich konventionell sind; nichts weist bei ihnen auf eine Auflehnung gegen die konventionellen Standards hin, die in ihrer Umgebung gelten; es ist unmöglich für sie, diese zu hinterfragen. Die Chance aber, daß ein ganzer Bereich konventioneller Standards untadelig wäre, ist gering. Der Sinn des Nachdenkens über hypothetische, vielleicht sehr unwahrscheinliche Situationen *scheint* schließlich darin zu liegen, daß man sich selbst oder einem anderen eine hypothetische Entscheidung für eine schlechte Handlung entlockt. Ich zweifle nicht, daß dies die Wirkung hat, Leute – die niemals in die Situation geraten werden, für die sie hypothetische Entscheidungen getroffen haben – dazu zu prädisponieren, in ähnlich schlechte Handlungen einzuwilligen oder diejenigen, die sie begehen, zu loben oder günstig darzustellen, solange es die Menge ebenfalls tut, auch wenn die vorgestellten hoffnungslosen Umstände nicht im geringsten gegeben sind.

(...)

Man kann unterscheiden – wenn ich einmal einfach anhand von Beispielen fortfahren darf – zwischen an sich ungerechten Handlungsweisen und solchen, die mit Rücksicht auf gegebene Umstände ungerecht sind. An sich ungerecht ist es, einen Menschen für etwas, das er offenkundig nicht getan hat, gerichtlich zu bestrafen. Natürlich kann so etwas geschehen, und es ist auf alle möglichen Weisen schon geschehen; durch Beste-

[6] Notwendigerweise ein seltener Fall: denn die positiven Vorschriften, z. B. „Ehre deine Eltern", geben praktisch nie eine direkte Handlungsanweisung und machen auch nur selten eine bestimmte Handlung notwendig.

chung falscher Zeugen, durch eine Gesetzesvorschrift, nach der etwas als Tatsache betrachtet wird, was anerkanntermaßen de facto nicht geschehen ist, durch offene Anmaßung von Richtern oder Machthabern, wenn sie mehr oder weniger unverblümt sagen: „Was kümmert es uns, daß du es nicht getan hast; wir sind entschlossen, dich dennoch dafür zu verurteilen." – Ungerecht unter gegebenen Umständen, z. B. unter normalen Umständen, ist es, jemandem sein vorgebliches Eigentum ohne legales Verfahren wegzunehmen, Schulden nicht zu bezahlen, Verträge nicht einzuhalten und viele andere Dinge dieser Art. Nun können aber die Umstände ganz klar eine sehr große Rolle spielen, wenn es darum geht, die Gerechtigkeit oder Ungerechtigkeit dieser oder ähnlicher Handlungsweisen zu beurteilen, und *manchmal* können diese Umstände auch erwartete Konsequenzen einschließen; z. B. kann der Anspruch eines Menschen auf ein bestimmtes Stück Eigentum nichtig werden, wenn sich durch dessen Beschlagnahme und Verwendung eine offensichtlich drohende Katastrophe verhindern läßt: etwa, wenn man mit einer ihm gehörenden Maschine eine Explosion hervorrufen könnte, bei der die Maschine zerstört würde, mit deren Hilfe aber eine Flut abgelenkt oder eine Kluft gesprengt werden könnte, die eine Feuersbrunst zum Stehen bringt. Nun bedeutet dies natürlich nicht, daß etwas, das normalerweise ungerecht, jedoch nicht an sich ungerecht ist, in jedem Fall durch eine einleuchtende Darstellung besserer Konsequenzen als gerecht hingestellt werden kann; keineswegs; aber die Probleme, die beim Versuch einer Grenzziehung (oder der Festlegung eines Grenzbereichs) auftreten würden, sind hier offensichtlich kompliziert. Und wenn hier auch sicher einige allgemeine Bemerkungen nötig und einige Grenzziehungen möglich sind, so wäre dennoch die Entscheidung in konkreten Fällen κατὰ τὸν ὀρθὸν λόγον, „nach dem, was vernünftig ist" zu treffen. Ob z. B. ein so und so langer Aufschub beim Begleichen der und der Schulden an eine Person in den und den Umständen von seiten eines in den und den Umständen befindlichen Schuldners ungerecht oder nicht ungerecht wäre, ist wirklich „nur nach dem, was vernünftig ist" zu entscheiden; hierfür kann es *prinzipiell* keinen anderen Maßstab geben, als daß man einige Beispiele anführt. Das bedeutet: Während es auf eine große Lücke in der Philosophie zurückzuführen ist, daß wir keine allgemeine Definition der Begriffe der Tugend und der Gerechtigkeit haben, sondern nur mit Beispielen operieren können, in denen wir diese Begriffe verwenden, gibt es einen Bereich, wo es nicht aufgrund irgendeiner Lückenhaftigkeit, sondern prinzipiell keine andere Art der Erläuterung gibt, als mit Hilfe von Beispielen: Und das ist dort, wo der Maßstab darin liegt, „was vernünftig ist", was natürlich *kein* Maßstab ist.

Weiter möchte ich nichts sagen über das, was in einigen Umständen gerecht, in anderen dagegen ungerecht ist, und über die Art und Weise, wie vorhergesehene Konsequenzen bei der Entscheidung über gerecht oder ungerecht eine Rolle spielen können. Zurück zu meinem Beispiel einer an

sich ungerechten Handlung: Die gerichtliche Bestrafung eines Menschen für etwas, wovon man als sicher annimmt, daß er es nicht getan hat, ist ungerecht; hierüber kann es absolut keine Meinungsverschiedenheiten geben. Nur solche Umstände oder vorhergesehene Konsequenzen, die die Beschreibung des Vorgangs als „gerichtliche Bestrafung eines Menschen für etwas, das er bekanntermaßen nicht getan hat", revidieren, können die Beschreibung dieses Vorgangs als ungerecht revidieren. Wer versuchen wollte, dies zu bestreiten, würde lediglich so tun, als wüßte er nicht, was „ungerecht" bedeutet: denn dies ist ein paradigmatischer Fall für Ungerechtigkeit.

(...)

Es ist der modernen Moralphilosophie – der Moralphilosophie aller namhaften englischen Ethiker seit Sidgwick – überlassen, Systeme zu konstruieren, nach denen ein Mensch eine tugendhafte Persönlichkeit sein *kann*, der sagt: „Wir brauchen das und das und werden es nur auf diese Weise erreichen"; d. h., es ist der Diskussion anheimgestellt, ob eine Handlungsweise wie die gerichtliche Bestrafung eines Unschuldigen nicht in bestimmten Umständen „richtig" sein könnte. Und obwohl die gegenwärtigen Oxforder Moralphilosophen einem *erlauben* würden, es „zu seinem Grundsatz zu machen", so etwas nicht zu tun, lehren sie doch eine Philosophie, wonach bei der Erörterung, was zu tun sei, die einzelnen Konsequenzen einer Handlung „moralisch" in Betracht gezogen werden *könnten*. Und falls diese so beschaffen wären, daß sie mit den „Zielen" des betreffenden Menschen kollidieren, könnte es einen Schritt in seiner moralischen Fortentwicklung bedeuten, wenn es ihm „gelänge" (um die Ausdrucksweise von Nowell-Smith zu gebrauchen)[7], die Handlung unter ein neu aufgestelltes Prinzip zu subsumieren; oder es könnte auch eine neue „Grundsatzentscheidung" sein, die ein Fortschritt in der Entwicklung seines moralischen Denkens wäre (um Hares Konzeption zu übernehmen), wenn er entschiede: Unter so und so gearteten Umständen sollte man die gerichtliche Verurteilung eines Unschuldigen erwirken. Und das kann ich nicht akzeptieren.

[7] P. H. Nowell-Smith, *Ethics*, Harmondsworth 1954.

Teil I: Elemente konsequentialistischer Rationalität

1. Kapitel: Konsequenzen

Ein vernünftiger Mensch bedenkt die *Folgen* seines Tuns. Falls er nicht gewiß sein kann, welche Folgen die eine oder andere Handlung hat, wird er nicht nur jeweils die wahrscheinlichste Folge einer Handlung, sondern auch weniger wahrscheinliche Folgen berücksichtigen: Er macht die Beurteilung der Handlung von der *Wahrscheinlichkeitsverteilung* ihrer Folgen abhängig. Nicht jede Bewertung von Handlungsfolgen ist mit der Vernünftigkeit der handelnden Person vereinbar. Eine vernünftige Person wird aber in jedem Falle eine Handlung wählen, die angesichts ihrer Folgen wünschenswerter erscheint als jede andere in der konkreten Situation mögliche (offenstehende) Handlung.

Das ist eine erste, grobe Charakterisierung dessen, was ich die *konsequentialistische Standardauffassung von Handlungsrationalität* nennen möchte. Sie liegt als Handlungsmodell verschiedenen sozialwissenschaftlichen Disziplinen, insbesondere den Wirtschaftswissenschaften[1], aber auch der sogenannten ökonomischen Theorie der Politik[2] zugrunde, und sie entspricht der gängigen Interpretation der rationalen Entscheidungs- und Spieltheorie[3]. Diese Standardauffassung ist mit einer Vielzahl von entscheidungs- und handlungstheoretischen Modellen vereinbar, auf die für die Zwecke dieser Untersuchung nicht im Detail eingegangen werden muß[4].

[1] Bei F. Y. Edgeworth ursprünglich auf der Annahme einer egoistischen Menschennatur gegründet, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, London 1881, wobei Edgeworth durchaus gewisse „Verunreinigungen" des reinen Egoismus durch zwischenmenschliche Sympathie einräumt. Die moderne ökonomische Gleichgewichtstheorie hat die Resultate, die Edgeworth aus egoistischem Verhalten der Akteure für die in Wettbewerbssituationen zu erreichenden „Optima" (die im wesentlichen dem Paretooptimum entsprachen) ableiten konnte, wesentlich verallgemeinert, ohne die Prämissen des „homo oeconomicus" aufzugeben; vgl. K. J. Arrow/F. H. Hahn: *General Competitive Analysis*, San Francisco 1971.

[2] Als Begründer dieses Forschungszweiges kann M. Olson jr. mit seinem Werk *The Logic of Collective Action*, Cambridge 1965 gelten, dt.: *Die Logik kollektiven Handelns*, Tübingen 1968. Vgl. a. G. Kirsch, *Ökonomische Theorie der Politik*, Tübingen 1974; B. Barry, *Neue politische Ökonomie*, Frankfurt/M. 1975 und als einführenden Überblick: F. Lehner, *Einführung in die neue Politische Ökonomie*, Königstein/Ts. 1981.

[3] Die nach wie vor beste Einführung in die Entscheidungs- und Spieltheorie bietet: R. D. Luce/H. Raiffa, *Games and Decisions*, New York/London/Sydney 1957; ohne die Behandlung von Interaktionssituationen: W. Stegmüller: *Entscheidungslogik (rationale Entscheidungstheorie)*, Berlin u. a. 1973.

[4] Entscheidungstheoretische Modelle unterscheiden sich von handlungstheoretischen insbe-

§ 1 Handlungskonsequenzen I: Handlungen

Es gibt unterschiedliche Arten von Konsequenzen. Wir beginnen mit einem besonderen Typ von Handlungskonsequenzen: Bisweilen sind *Konsequenzen von Handlungen andere Handlungen.* Die folgenden Beispiele zeigen, daß dabei verschiedene Fälle zu unterscheiden sind.

1) Der Polizist hebt den Arm und stoppt damit die Autokolonne.

2) Der Polizist entsichert die Pistole, betätigt den Abzug und tötet damit den Kidnapper.

3a) Hans nimmt, ohne die anderen Gäste zu fragen, das letzte Stück Kuchen und bringt damit den Gastgeber in Verlegenheit.

3b) Hans geht heute abend allein ins Kino und bricht damit sein Versprechen, heute mit Lisa auszugehen.

3c) Hans erzählt seinen Freunden von Peters Vergangenheit und rächt sich damit an Peter.

Im ersten Beispiel spielt offensichtlich die Konvention, aufgrund derer Polizisten den Verkehr regeln, eine wesentliche Rolle. Ohne diese Konvention könnte der Polizist den Verkehr nicht dadurch stoppen, daß er den Arm hebt. Möglicherweise könnte er ihn dadurch stoppen, daß er sich in den Weg stellt. Im zweiten Beispielt tötet der Polizist den Kidnapper dadurch, daß er den Abzug betätigt – und das geschieht unabhängig von jeder Konvention. In beiden Fällen sind die Handlungen als Konsequenzen nicht zwingend: Es könnte sein, daß die Autokolonne nicht stehen bleibt, und es könnte sein, daß die Kugel den Kidnapper verfehlt. Es lassen sich also offensichtlich zwei Fälle von Handlungen als Konsequenzen von Handlungen unterscheiden. Wir nennen den ersten Fall: *„Das Hervorbringen einer Handlung durch eine Handlung qua Konvention"* und den zweiten Fall: *„Das kausale Hervorbringen einer Handlung durch eine Handlung"*.

Die Fälle 3a, 3b und 3c zeigen, daß diese Fallunterscheidung nicht so eindeutig ist, wie man zunächst annehmen könnte. Kommt der Gastgeber nur aufgrund einer Konvention in Verlegenheit? Oder verursacht das Ver-

Fortsetzung Fußnote von Seite 11

sondere dadurch, daß sie den Handlungsbegriff selbst unexpliziert lassen. Insofern, könnte man sagen, kommt der Handlungstheorie ein logisches „Prius" zu: Die Entscheidungstheorie gewinnt erst im Kontext handlungstheoretischer Klärungen die wünschenswerte Explizitheit. Zum Problem der Explizitheit entscheidungstheoretischer Modelle vgl. W. Spohn, *Grundmodelle der Entscheidungstheorie*, Kronberg/Ts. 1978. Andererseits kann man in jüngster Zeit eine stärkere Vernetzung dieser beiden Disziplinen beobachten. Zu nennen ist in diesem Zusammenhang das Projekt einer integrierten Handlungstheorie von H. Lenk. Wie fruchtbar eine interdisziplinäre Kooperation auf diesem Gebiet sein kann, zeigen die daraus hervorgegangenen Sammelbände *Handlungstheorie – interdisziplinär*. Bd. I–IV, München 1977–84. Als Dokument dieser stärkeren Vernetzung kann auch der interessante Ansatz einer generellen Handlungslogik von Aqvist gelten, vgl. L. Aqvist, „A New Approach to the Logical Theory of Action and Causality", in: *Logical Theory and Semantic Analysis*, hg. von S. Stenlund, Dordrecht 1974, S. 73–91.

halten von Hans bei jedem Gastgeber, unabhängig von einer sozialen Regelverletzung, Verlegenheit? Ist das „In-Verlegenheit-Bringen" eine Handlung von Hans? Dies scheint nur dann der Fall zu sein, wenn Hans das „In-Verlegenheit-Bringen" intendiert hat. Wenn dem nicht so ist, ist es für den normalen Sprachgebrauch zumindest fraglich, ob das „In-Verlegenheit-Bringen" als Handlung gelten kann – eine Konsequenz einer seiner Handlungen ist es in jedem Fall.

3b ist nicht einfach ein weiteres Beispiel wie 2; denn es gibt keine Konvention, die sich auf die Handlung „allein ins Kino gehen" bezieht. Aber „Versprechen geben" ist regelgeleitet, und diese Regeln verlangen von Hans, daß er an diesem Tag nicht allein ins Kino geht. 3c hingegen scheint eher dem Beispiel 2 als dem Beispiel 1 zu ähneln: Hans – so könnte man paraphrasieren – verletzt Peter seelisch, indem er aus der Vergangenheit des Freundes erzählt. Die seelische Verletzung ist eine kausale Folge der Erzählung. Der Einwand, was als Rache gelten kann, sei konventionsgeleitet, ist irreführend, da es nicht um (möglicherweise konventionelle[5]) Regeln des Sprachgebrauchs geht[6], sondern darum, ob sich die Handlungskonsequenz, sich gerächt zu haben, aufgrund einer Konvention ergibt (was hier nicht der Fall zu sein scheint).

Obwohl diese Unterscheidung[7] (Hervorbringen einer Handlung durch eine Handlung qua Konvention vs. kausales Hervorbringen einer Handlung durch eine Handlung) – wie die Beispiele gezeigt haben – nicht ganz trennscharf ist, markiert sie zwei Arten des Hervorbringens einer Handlung durch eine Handlung, die in der weiteren Analyse unterschiedlich zu behandeln sind. Wir wollen von „Hervorbringung" einer Handlung h durch eine Handlung h' nur dann sprechen, wenn h eine andere Handlung ist als h'. Das Beispeil von G. E. M. Anscombe[8]: jemand bewegt seinen Arm, damit betätigt er eine Pumpe, damit füllt er ein Wasserreservoir auf und damit vergiftet er die Hausbewohner, ist kein Beispiel für unterschiedliche Beschreibungen einer identischen Handlung[9]. Wenn alle der

[5] Vgl. D. Lewis, *Convention: A Philosophical Study*, Cambridge 1969; dt. *Konventionen. Eine sprachphilosophische Abhandlung*, Berlin/New York 1975; G. H. Bird, „Intentions and Conventions", in: *Logique et Analyse*, Nouvelle Serie 67/68 (1974) 495–507; A. Kemmerling, *Konvention und sprachliche Kommunikation* (Diss.), München 1975; G. Meggle: *Handlungstheoretische Semantik*, Berlin/New York 1980.

[6] Es gibt auch für „töten" Regeln des korrekten Sprachgebrauchs.

[7] Für A. Goldman gibt es neben „causal generation" und „conventional generation" noch „simple generation" und „augmentation", vgl. *A Theory of Human Action*, Englewood Cliffs 1970. Tatsächlich läßt sich die dritte je nach Situation auf die erste oder zweite zurückführen, und die vierte gehört zu einer anderen Thematik, nämlich derjenigen der Handlungsidentität. Vgl. dazu auch: R. Tuomela, *A Theory of Social Action*, Dordrecht 1984, S. 161 f.

[8] *Intention*, Oxford 1957, S. 23–26.

[9] Zum Problem der Handlungsidentität vgl.: A Cody, „Can A Single Action Have Many Different Descriptions?", *Inquiry* 10 (1967) 164–180; D. Davidson, „The Individuation of Events" (1969), in: ders., *Essays on Actions and Events*, Oxford 1980; A. Goldman, „The Individuation of Events", *J. Phil.* 68 (1971) 761–774, dt. in: *Analytische Handlungstheorie*

vier genannten Vorkommnisse tatsächlich Handlungen beschreiben, dann sind es vier unterschiedliche Handlungen – das zeigt sich schon an der Asymetrie der „Relation des Hervorbringens": Durch die Auffüllung des Reservoirs vergiftet die Person die Hausbewohner, aber nicht umgekehrt. Ob allen vier Vorkommnissen Handlungen der Person entsprechen, scheint davon abzuhängen, was die Person weiß (wenn sie nicht weiß, daß das Wasser vergiftet ist, entspricht der Vergiftung sicher keine Handlung des Vergiftens) und was die Person intendiert. Für eine rationale Person gibt es natürlich einen engen Zusammenhang zwischen ihrem Wissen und ihren Intentionen: wenn sie weiß, daß das Wasser vergiftet ist, und wenn sie weiß, daß die Betätigung der Pumpe das Wasser in ein Reservoir leitet, von dem die Hausbewohner ihr Trinkwasser schöpfen, dann kann kaum angenommen werden, daß die Person nicht auch eine Intention hatte, die es rechtfertigt, ihr die Handlung des Vergiftens zuzuschreiben.

Die Kategorie des konventionellen Hervorbringens einer Handlung durch eine Handlung darf nicht mit einer *logischen Beziehung zwischen Handlungen* verwechselt werden. Bisweilen werden spezifischere Handlungen als logisch notwendige Konsequenzen von Handlungen hervorgebracht[10]. Nicht nur Handlungsfolgen, die selbst Handlungen sind, können den Charakter *logischer Notwendigkeit* haben. Wenn eine Person eine Tür zum Zeitpunkt t abschließt, ist der Sachverhalt, daß diese Tür zu einem Zeitpunkt t'[11] abgeschlossen ist, eine logisch-notwendige Folge. Wir nennen eine logisch notwendige Folge einer Handlung h ein *Ergebnis von h*[12]. Ein Ergebnis einer Handlung h einer Person i kann selbst wieder eine Handlung von i sein. Im allgemeinen sind Ergebnisse von Handlungen jedoch *Zustände* oder *Ereignisse,* die keine Handlungen sind. Der Polizist vollzieht die Handlung, den Verkehr zu stoppen (im perlokutionären Sinne) nur, wenn die Wagenkolonne kurz darauf anhält (Ereignis) bzw. wenn kurz darauf die Wagenkolonne steht (Zustand). Hans schließt die die Tür nur ab, wenn dabei der Riegel einrastet (Ereignis) bzw. wenn die Tür anschließend abgeschlossen ist (Zustand)[13].

Fortsetzung Fußnote von Seite 13

Bd. I: Handlungsbeschreibungen, hg. von G. Meggle, Frankfurt/M. 1985, S. 332–353; M. Bradsley, „Actions and Events: The Problem of Individuation", *Am.Phil.Quart.* 12 (1975) 263–276.

[10] „Logisch notwendig" soll hier nichts anderes heißen als „unabhängig von (empirisch) kontingenten Sachverhalten (einschließlich bestehender Konventionen) notwendig".

[11] t' größer t – für welches Zeitintervall das gelten muß (es könnte ja sein, daß sich die Tür aufgrund eines Defektes wieder von selbst öffnet) und wie groß der Abstand von t' und t sein darf, ist für das Beispiel nicht wesentlich.

[12] In Anlehnunung an G. H. v. Wright, *Explanation and Understanding,* Ithaca 1971 und Tuomela (1984), Kap. 3, II.

[13] Zahlreiche Ausdrücke haben eine Doppelbedeutung besonderer Art: „Beweis" ist einmal der *„gelungene Beweis"* und in „dieser Beweis ist nicht korrekt" auch der „Beweisversuch" – eine ähnliche Doppelbedeutung gibt es für Handlungen: „Er schließt die Türe ab" (ohne Erfolg) impliziert nicht, daß die Türe dann auch abgeschlossen ist.

§ 2 *Atom-, Basis- und Abschlußhandlungen*

In der Handlungstheorie wird weithin zwischen Basishandlungen und Nicht-Basishandlungen unterschieden. Für Basishandlungen sind verschiedene Kriterien diskutiert worden[14]. Im folgenden werden Handlungen, die nicht Konsequenz anderer Handlungen sind, als *Basishandlung* bezeichnet. Die ersten Beiträge zur Theorie der Basishandlung identifizierten *komplexe* und *Nicht*-Basishandlungen sowie *einfache* und *Basis*handlungen. Als typisches Beispiel für Basishandlungen galten einfache Körperbewegungen. Danto sprach von einem Repertoire an Basishandlungen, das eine Person habe und das einem bestimmten Repertoire an Körperbewegungen entspreche[15]. Aber dabei wurden zwei Merkmale von Handlungen vermengt: ihre Komplexität als raum-zeitlicher Vorgang und ihr Verhältnis zu anderen Handlungen.

Es mag sein, daß bestimmte einfache Körperbewegungen unter normalen Bedingungen nicht mehr in Teilhandlungen zerlegbar sind. Wenn ich meinen rechten Arm hebe, vollziehe ich normalerweise nicht eine ganze Folge von Handlungen: den Arm die ersten zehn Zentimeter heben, die zweiten zehn Zentimeter heben etc. Vielleicht ist der Mensch so beschaffen, daß er bestimmte Körperbewegungen unter normalen Bedingungen nur als Basishandlungen im oben definierten Sinne vollziehen kann: Er vollzieht keine anderen Handlungen, die diese Handlung hervorbringen. Aber ob dem so ist, ist nicht wesentlich für den Begriff der Basishandlung. Eine bestimmte singuläre Handlung h ist dann eine Basishandlung, wennn keine anderen Handlungen vollzogen worden sind, die h zur Konsequenz haben. Bestimmte Handlungen können also auch dann Basishandlungen sein, wenn sich die Theorie eines Repertoires an Körperbewegungen als Basishandlungen als falsch herausstellen sollte. Basishandlungen müssen keine einfachen Körperbewegungen sein. Auch Handlungen, die sich in sehr komplexen Vorgängen manifestieren, können Basishandlungen sein. Was für einen geübten Autofahrer eine Basishandlung ist – etwa rechts abbiegen – mag für den Anfänger eine Nicht-Basishandlung sein, d. h. eine Handlung, die Ergebnis einer Reihe von anderen Handlungen ist: die rechte Hand nach links oben versetzen, den Blinker betätigen etc.

[14] Vgl. A. I. Melden, *Free Action*, London 1961, bes. Kap. 6. Explizit von Basishandlungen spricht m. W. als erster A. C. Danto, „What We Can Do", *J. Phil.* 60 (1963) 435–445. Vgl. a. A. C. Danto, „Basic Actions", *Am.Phil.Quart.* 2 (1965) 141–148; F. Stoutland, „Basic Actions and Causality", *J. Phil.* 65 (1968) 467–475; M. Brand, „Danto on Basic Actions, *Nous* 2 (1968) 187–190; J. R. Martin, „Basic Actions and Simple Actions", *Am.Phil.Quart.* 9 (1972) 59–68. Auch R. M. Chisholm stützt sich auf diesen Begriff: „The Descriptive Element in the Concept of Action", *J. Phil.* 61 (1964) 613–624, ebenso R. Taylor, „Thought and Purpose", in: *The Nature of Human Action*, hg. von M. Brand, Glenview 1970, S. 267–282, bes. S. 276f. Die Artikel von Danto (1965) und Martin (1972) sind zusammen mit einer Kritik der Theorie der Basishandlungen von A. Baier, „The Search for Basis Actions", *Am.Phil.Quart.* 8 (1971) 161–170, dt. in: G. Meggle (1985) erschienen.

[15] Vgl. A. C. Danto, „Basic Actions", *Am.Phil.Quart.* 2 (1965) 141–148.

Eine Handlung kann *Ergebnis* einer Konjunktion von Handlungen sein, aber sie ist dann selbstverständlich nicht *Ergebnis* jeder einzelnen Handlung dieser Konjunktion. Wenn eine Handlung h Ergebnis einer Konjunktion von Handlungen h_1 & h_2 & ... & h_n ist, so ist es *logisch* ausgeschlossen, daß h_1 und h_2 und ... und h_3 vollzogen wurden, ohne daß h vollzogen wurde. Basishandlungen sind Handlungen, die nicht Ergebnis oder Konsequenz anderer Handlungen der betreffenden Person sind.

Abschlußhandlungen nennen wir Handlungen, die keine anderen Handlungen mehr als Ergebnis oder Konsequenz haben. Der Polizist, der seinen Arm hebt, vollzieht damit eine Basishandlung: Er tut sonst nichts, was als Ergebnis oder Konsequenz die Handlung des Armhebens nach sich ziehen würde. Er hebt einfach den Arm. Sein Armheben ist keine Abschlußhandlung; denn sie hat eine weitere Handlung – nämlich die Anweisung an die Autofahrer stehenzubleiben – als Konsequenz (nicht jedoch als Ergebnis). Es ist natürlich nicht ausgeschlossen, daß eine Handlung sowohl Basis- wie Abschlußhandlung ist. Sollte diese Handlung eine Massenkarambolage zur Folge haben, ist dies keine Handlung des Polizisten, sondern die Anweisung stehenzubleiben scheint in diesem Fall eine Abschlußhandlung zu sein.

Die Handlung des Armhebens ist ein recht komplexer Vorgang und man kann ihn sich aus einzelnen Vorgängen zusammengesetzt denken, die allesamt *potentielle* Handlungen des Polizisten sind. Wenn er eine Armverletzung hat, könnten einzelne dieser Vorgänge – etwa anfangs eine leichte Abwinkelung des Unterarms – zu Handlungen werden. Es kann hier offenbleiben, ob es *einfachste* Handlungen gibt, die sich nicht mehr aus Vorgängen zusammensetzen, sondern jeweils einzeln potentielle Handlungen sind. Das ist in erster Linie eine Frage der Psychologie und Physiologie. Wenn dem so wäre, gäbe es einfachste Handlungen, die zugleich Basishandlungen wären. Das Dantosche „Repertoire" wäre dann aber nicht das Repertoire an Basishandlungen generell, sondern das Repertoire einer (vergleichsweise kleinen) Untermenge von Basishandlungen. Wir wollen es das *Repertoire der Atomhandlungen* nennen. Sollte es keine Grenze der Zerlegbarkeit geben, so gibt es nach wie vor singuläre Handlungen die Basishandlungen, aber keine Handlungstypen, die Atomhandlungen sind.

§ 3 *Savage-Konsequenzen*

Handlungen können, wie wir gesehen haben, andere Handlungen „hervorbringen": Ein Typ von Handlungskonsequenzen sind Handlungen. Um den Konsequentialismus von anderen Theorien abgrenzen zu können, ist es notwendig, diesen Typ Handlungskonsequenzen von jenem zu unterscheiden, der für die Bestimmung konsequentialistischer Rationalität maßgebend ist. Der Begründer der modernen rationalen Entscheidungs-

theorie, L. J. Savage, hat ein formales Modell entwickelt, anhand dessen sich diese Unterscheidung präzisieren läßt.

Savage möchte in seiner Grundlegung der Statistik[16] die Rolle der Vernunft bei der Entscheidungsfindung klären[17]. Zu diesem Zweck werden folgende Begriffe – zunächst weitgehend ohne intuitive Vorklärungen – eingeführt: Die *„Welt"* als das Objekt, das für die (entscheidende) Person relevant ist; der *„Weltzustand"* als eine Beschreibung der Welt, die keinen relevanten Aspekt ausläßt; den *„tatsächlichen Weltzustand"* als derjenige Weltzustand, der tatsächlich besteht, d. h. eine wahre Beschreibung der Welt; das *„Ereignis"* als eine Menge von Weltzuständen[18].

Diese begrifflichen Festsetzungen verlieren ihre Willkürlichkeit erst vor dem Hintergrund der voll entwickelten Theorie. So läßt sich z. B. zeigen, daß die „Vergrößerung" der in der Entscheidungssituation berücksichtigten Welt nicht zu Inkonsistenzen führt, so daß es naheliegend ist, die „größte" Welt zu berücksichtigen: die vollständige vergangene, gegenwärtige und zukünftige Geschichte des Universums. *Konsequenzen* werden als „Zustände der Person" eingeführt und Zuständen der Welt gegenübergestellt. *Handlungen* können dann mit den (möglichen) Konsequenzen identifiziert[19] werden: Eine Handlung ist eine Funktion, die jedem Weltzustand eine Konsequenz zuordnet.

Das Modell von Savage sei an einem alltäglichen Beispiel einer Entscheidungssituation erläutert: Ich stehe heute vor der Frage, mit welchem Verkehrsmittel ich morgen von München nach Frankfurt reise (wir wollen vereinfachend annehmen, daß ich aus irgendeinem Grund gezwungen bin, meine Entscheidung heute zu treffen, und daß die getroffene Entscheidung weder vom Handelnden rückgängig gemacht noch durch äußere Umstände unwirksam werden kann). Es stehen mir drei Alternativen offen: Nach Frankfurt mit dem eigenen Wagen, der Bahn oder dem Flugzeug zu reisen. Meine subjektiven Präferenzen sind je nach den auftretenden Umständen unterschiedlich. So ziehe ich etwa bei gutem Wetter wegen der Bequemlichkeit und Schnelligkeit den Flug vor, bei dichtem Schneetreiben die Bahn (denn dann besteht die Gefahr, daß die Maschine

[16] L. J. Savage, *The Foundations of Statistics*, New York ²1971.

[17] „It may be said to be the purpose of this book, and indeed of statistics generally, to discuss the implications of reasoning for the making of decisions." Man könnte daher von einem intellektualistischen Ansatz Savages sprechen, der ausschließt, daß Tiere rational handeln, da sie zur Folgenabwägung kaum in der Lage sind – Savage spricht hier sogar von „formal reasoning", das Menschen vorbehalten sei, und zwar in einem geringeren Umfang als man es sich wünschen mag (Savage (1971, S. 6). Diese Sicht ist nicht selbstverständlich, denn der Abwägungsprozeß selbst geht als Bedingung rationaler Entscheidung an keiner Stelle in die formale Theorie rationaler Entscheidung ein. Präferenzverhalten kennt man jedoch auch von höheren Tieren. Es ist nicht ausgeschlossen, daß sich aufgrund des manifesten Präferenzverhaltens eines Tieres eine konsistente subjektive Bewertungsfunktion bilden läßt – so daß erst zu zeigen wäre, inwiefern ein solches tierisches Verhalten im Sinne der Savageschen Theorie nicht rational ist.

[18] Vgl. Savage (1971), S. 14ff.

[19] Vgl. Savage (1971), S. 14ff.

in Frankfurt nicht landen kann, und Autofahren bei Schneetreiben ist mühselig und gefährlich) und bei stürmischem Regenwetter das Auto, denn bei Sturm habe ich Flugangst, und andererseits behindern mich diese Wetterbedingungen nicht so sehr, daß ich die Nachteile der Bahnfahrt in Kauf nehmen möchte.

Ich bin mir dieser und anderer[20] (bedingter) Präferenzen sicher. Zu ihrer Bestimmung benötige ich nicht die Hilfe einer Theorie. Aber selbst wenn diese Präferenzen eindeutig sind und ich alle relevanten Umstände in Betracht gezogen habe, könnte ich unsicher sein, welche Entscheidung für mich die beste ist. Es scheinen nur zwei Aspekte der Situation von Belang: die Folgen der Entscheidungsalternativen und ihre durch die getroffene Entscheidung bedingten Wahrscheinlichkeiten. Dieses Vorverständnis von (konsequentialistischer) Handlungsrationalität hat also offensichtlich zur Voraussetzung, daß die (subjektiven) Werte der möglichen Handlungs*konsequenzen* vergleichbar sind und (von mir) beurteilt werden können – unabhängig von meiner Neigung, die eine oder andere *Handlung* vorzuziehen.

In unserem Beispiel würde man als Konsequenzen der jeweiligen Entscheidung bestimmte Ergebnisse oder Vorgänge ansehen: Wenn ich mich entscheide zu fliegen, sind die möglichen Konsequenzen etwa die folgenden: bei stürmischem Wetter unruhiger Flug, unsanfte Landung, Flugangst; bei Nebel Landung auf einem anderen Flughafen mit entsprechenden Verzögerungen; unter allen anderen Wetterbedingungen: rasche und bequeme Reise. Daß ein bestimmtes Wetter herrscht, entspricht einem Weltzustand oder, bei der Wahl einer „größeren" Welt, einem Ereignis, also einer Menge von Weltzuständen. Dieser Schilderung möglicher Handlungsfolgen liegen bestimmte (unausgesprochene) Annahmen über empirische Bedingungen zugrunde. Natürlich wäre es für die Beurteilung der Situation relevant, wenn etwa auch Vulkanausbrüche während des Fluges nicht ausgeschlossen werden könnten oder wenn der Handelnde annehmen müßte, daß seine Flugangst eventuell Spätfolgen in Form neurotischer Störungen nach sich ziehen würde. Es gibt weder zeitliche noch räumliche Beschränkungen, die bestimmte Umstände oder (allgemeiner) Ereignisse als irrelevant für eine Entscheidungssituation ausschließen. Relevant ist das, was die Handlungswahl unter dem gewählten Rationalitätskriterium beeinflussen könnte. Insofern setzt die modellhafte Darstellung der Entscheidungssituation implizit immer schon die Anwendung des Rationalitätskriteriums voraus[21]. Wenn man in dem gewählten Beispiel aus-

[20] Die genannten Umstände sind nicht die einzigen, die relevant sind.

[21] Dieser auftretende Zirkel ist nicht nur hier, sondern auch für naturwissenschaftliche Theorien einer gewissen Allgemeinheitsstufe unumgänglich. So gilt das zweite Newtonsche Gesetz eben nur in Inertialsystemen, und was ein Inertialsystem ist, läßt sich wiederum nur mit Hilfe der Newtonschen Gesetze formulieren. Dieser Anwendungszirkel macht die jeweilige Theorie jedoch nicht gegenüber Widerlegungsversuchen grundsätzlich immun: Die

schließlich das Wetter während der Reise für relevant hält[22], so ist die Auswahl der Umstände davon bestimmt, welchen Wetterzuständen (allgemein: Weltzuständen) jeweils eindeutige Konsequenzen (im Sinne Savages: subjektive Zustände der Person) entsprechen. Die Einteilung Regen/Nicht-Regen etwa würde jeweils Wetterzustände umfassen, die für das Befinden der Person ganz unterschiedliche Folgen hätten; denn Regen ist mit schwerem Sturm und Windstille und Nicht-Regen sowohl mit Schneetreiben wie mit Sonnenschein vereinbar, was in unserem Beispiel Auswirkungen auf die Handlungspräferenzen hätte. Wesentlich für die Abschätzung der Handlungsfolgen ist, daß die Umstände des Modells handlungsunabhängig und so gewählt sind, daß sie sich wechselseitige ausschließen[23].

Konsequenzen werden von Savage als „Zustände der Person" charakterisiert. Die in der Konsequenzenmatrix aufgeführten Ereignisse und Vorgänge sind daher als Kennzeichnungen bestimmter Zustände der handelnden Person aufzufassen. Diese Interpretation verleiht dem Modell seinen konsequentialistischen Charakter: Rational ist diejenige Handlung, die den „Zustand der handelnden Person" optimiert. Die Gegenstände der subjektiven Wertfunktion sind also nicht beliebig. So ist z. B. die Übereinstimmung der infragestehenden Handlung mit einem Befehl oder einer ethischen Handlungsregel nicht Element der Konsequenzenmatrix und damit auch nicht Gegenstand der subjektiven Bewertung.

In der Anwendung des Modells muß man sich immer auf eine „kleine Welt" beschränken. Eine vollständige Auflistung aller relevanten Merkmale der Entscheidungssituation ist unmöglich. In der Anwendung wird daher implizit freigebig von Hintergrundwissen Gebrauch gemacht, so daß man sich im wesentlichen darauf beschränken kann, Ereignisse und Vorgänge zu erwähnen, die zwar für sich genommen für eine rationale Beurteilung in keiner Weise ausreichend sind, die aber, als Hinweise auf komplexere Ereignisstrukturen verstanden, eine angemessene Repräsentation der Situation darstellen.

Nicht alle Entscheidungssituationen lassen sich so modellieren, daß dies Forderung erfüllt ist. W. Spohn bringt folgendes Gegenbeispiel: Eine Person hat in einer Klausurprüfung eines von zwei vorgelegten Themen zu bearbeiten. Wenn es ihr nur auf das Bestehen der Prüfung ankommt, wird sie das Thema wählen, von dem sie mit höherer Wahrscheinlichkeit an-

Fortsetzung Fußnote von Seite 18

T-Theoretizität der Begriffe, mit deren Hilfe eine gesetzesartige Hypothese formuliert ist, schließt ihre Widerlegung nicht aus. Vgl. dazu: J. D. Sneed, *The Logical Structure of Mathematical Physics*, Dordrecht 1971 sowie W. Stegmüller, *Theorie und Erfahrung. Die Entwicklung des Strukturalismus seit 1973*, Berlin 1986.

[22] Ob das zutrifft, ist, wie sich aus dem Zirkelargument ergibt, eine empirische Frage.

[23] Wenn die relevanten Ereignisse $e_1, e_2, \ldots, e_n$ sind, läßt sich diese Forderung leicht durch die Auflistung aller Normalformen auf der Menge der diese Ereignisse beschreibenden Sätze darstellen.

nehmen kann, die Prüfung zu bestehen[24]. In diesem Beispiel macht es Schwierigkeiten, Umstände, für die der Handelnde subjektive handlungsunabhängige Wahrscheinlichkeiten hat, und Konsequenzen, die er unabhängig von seinen Erwartungen zukünftiger Ereignisse bewertet, zu bestimmen. Das einfache, intuitiv überzeugende Grundmodell von Savage scheint hier bereits überfordert zu sein[25]. Die beiden für die Explikation konsequentialistischer Handlungsrationalität wesentlichen Probleme, sowohl das der Wahl einer angemessenen Welt im Sinne von Savage als auch das der Beziehung von Umständen und Konsequenzen, lassen sich jedoch im Rahmen einer „modell-platonistischen" Betrachtungsweise klären.[26]

§ 4 *Handlungskonsequenzen II: Weltverläufe*

Aus realistischer Sicht[27] ist das Universum zu jedem Zeitpunkt in einem ganz bestimmten *Zustand*. Einzelne Personen kennen immer nur einige wenige Merkmale dieses Zustandes. Aber auch das kumulative Wissen der Bürokratien und Wissenschaftler-Gemeinschaften beschränkt sich auf einen winzigen Ausschnitt alles Wissensmöglichen bezüglich des jeweiligen Zustandes. In jedem Fall läßt sich ein Weltzustand schon aus praktischen Gründen nicht vollständig beschreiben. Was eine korrekte *Beschreibung* (von „Ausschnitten") eines Zustandes des Universums ist, richtet sich selbstverständlich nach den Regeln und Begriffen der verwendeten Sprache. Aber der Zustand selbst ist sprach- und theorieunabhängig, ihm entsprechen in unserem Modell real existierende Entitäten und damit stellt sich die Frage nach der Ontologie dieses Modells.

[24] W. Spohn, *Grundmodelle der Entscheidungstheorie*, Kronberg/Ts. 1978, S. 50.

[25] Andere entscheidungstheoretische Modelle verzichten wegen solcher Komplikationen auf die gesonderte Berücksichtigung von Umständen, was jedoch die Interpretation der Grundbegriffe eher erschwert, da nun handlungsunabhängige Ereignisse zu Bestandteilen der Konsequenzen werden. Wir führen daher die Diskussion entscheidungstheoretischer Modelle nicht fort, sondern greifen im folgenden auf ein Modell der philosophischen Handlungstheorie zurück.

[26] Von den Modellen intensionaler Semantik haben im Rahmen der Handlungstheorie insbesondere G. H. v. Wright, L. Aqvist und F. v. Kutschera überzeugend Gebrauch gemacht, vgl. G. H. v. Wright, *Erklären und Verstehen*, Königstein/Ts. 1974, ²1984; L. Aqvist, „A New Approach to the Logical Theory of Action and Causality", in: *Logical Theory and Semantic Analysis*, hg. von S. Stenlund, Dordrecht 1974, S. 73–91. Eine modifizierte und ausgebaute Version dieses Modells hat F. v. Kutschera entwickelt: „Grundbegriffe der Handlungslogik", in: H. Lenk (1980) S. 67–106 – eine gute Einführung zur intensionalen Semantik bietet F. v. Kutschera, *Einführung in die intensionale Semantik*, Berlin 1976.

[27] Wir stellen uns hier auf den Standpunkt eines metaphysischen Realismus. Das ist aber für das Argument nicht wesentlich. Vielmehr dient es einer naheliegenden Interpretation der einzuführenden Grundbegriffe, die unter anderen ontologischen und erkenntnistheoretischen Annahmen ebenfalls sinnvoll sind, aber dann in einem anderen und z. T. wesentlich komplizierteren Sinne zu interpretieren wären. Insbesondere würden sie sich problemlos auch im Kontext eines „internen Realismus", wie Putnam seine pragmatische Version von Realismus nennt, interpretieren lassen, vgl. Putnams Rede „Realism and Reason", abgedruckt in: H. Putnam, *Meaning and the Moral Sciences*, London 1978 und W. Stegmüller, *Hauptströmungen der Gegenwartsphilosophie*, Bd. II, Stuttgart 1979, S. 446 ff.

Grundlegend ist dafür der Begriff eines *generischen Sachverhaltes*[28]. Beispiele für generische Sachverhalte sind etwa, daß es regnet oder daß eine bestimmte Person Minister ist etc. Ein generischer unterscheidet sich von einem singulären Sachverhalt dadurch, daß ein generischer Sachverhalt zu verschiedenen Zeitpunkten gegeben sein kann, was aus begrifflichen Gründen bei einem singulären Sachverhalt ausgeschlossen ist. Ein generischer Sachverhalt kann im Gegensatz zu einem singulären in verschiedenen Weltzuständen vorligen. Dabei ist es eine bloße Frage der Zweckmäßigkeit im Sinne einer möglichst eindeutigen und intuitiv nachvollziehbaren Interpretation der verwendeten Grundbegriffe, ob man den Begriff des (generischen) Sachverhaltes mit Hilfe des Begriffes „Weltzustand" einführt oder umgekehrt. Im ersten Fall sind (generische) Sachverhalte Mengen von Weltzuständen, und im zweiten sind Weltzustände vollständige Kombinationen generischer Sachverhalte.

Angenommen, es gebe n logisch voneinander unabhängige generische Sachverhalte und diesen Sachverhalten entsprechen n Sätze, die diese Sachverhalte beschreiben, dann ist eine vollständige Beschreibung eines Weltzustandes eine Konjunktion mit n Gliedern. Jedes dieser Glieder ist entweder einer dieser n Sätze oder eine Negation einer dieser n Sätze, wobei jeder Satz in der Konjunktion genau einmal vorkommt (konjunktives Fundamentalschema). Bei n zugrundegelegten Sachverhalten ergeben sich 2^n mögliche Beschreibungen unterschiedlicher Weltzustände, vorausgesetzt, jeder Sachverhalt läßt sich eindeutig durch einen Satz beschreiben. Wenn jeweils mehrere Sätze einen Sachverhalt beschreiben, nennen wir diese Sätze „(sachverhalts-)äquivalent", damit gibt es 2^n Äquivalenzklassen bzw. mehrere Mengen untereinander nicht sachverhaltsäquivalenter Sätze mit jeweils 2^n Elementen.

Sachverhalte einer Menge M sind logisch unabhängig, wenn jedes konjunktive Normalschema der sie beschreibenden Sätze möglich ist. Sachverhalte aus M sind kausal unabhängig, wenn jeder durch eine Normalform beschriebene Zustand mit den bestehenden kausalen Abhängigkeiten vereinbar ist. Ein Weltzustand bzw. für die reale Welt wahre Sachverhalte, also Tatsachen, besteht in diesem Modell aus Sachverhalten. Sachverhalte sind die ontologischen Bausteine dieses Konzeptes, weder Eigenschaften noch Einzeldinge werden als real existierend vorausgesetzt – das hat es mit der Wittgensteinschen Welt des *Tractatus* gemeinsam[29].

Gemäß dieser Charakterisierung durch generische Sachverhalte ist es nicht ausgeschlossen, daß der Zustand der Welt zu zwei unterschiedlichen Zeitpunkten gleich ist: Der Zeitpunkt ist keine Eigenschaft des Zustandes. Der logische Formalismus des Modells kann sich auf die zweiwertige Aussagenlogik, ergänzt durch eine (diskrete) Zeitlogik, beschränken.

[28] Dieser Begriff des generischen Sachverhaltes stützt sich auf G. H. v. Wright, *Norm an Action*, London 1963, bes. Kap. 2.

[29] L. Wittgenstein, *Tractatus logico-philosophicus* (1921), in: *Schriften 1*, Frankfurt/Main 1960.

Ein *Weltverlauf* ist eine Folge von Weltzuständen $z_1, z_2, \ldots, z_r$. Bei n generischen Sachverhalten und einer Weltgeschichte der Länge r (r aufeinanderfolgende Weltzustände) ergeben sich 2^{rn} mögliche Weltverläufe. Dabei stellt für handlungstheoretische Zwecke die Frage der Existenz eines ersten Weltzustandes kein Problem dar. Wesentlich ist, daß es von der Entstehung der Welt bis zum gegenwärtigen Zeitpunkt einen eindeutigen Weltverlauf im Sinne einer direkten Aufeinanderfolge von Weltzuständen gibt.

Auf jeder Stufe i des Weltverlaufes realisiert sich genau ein nachfolgender Zustand z_{i+1}. Wenn der Weltverlauf nicht vollständig determiniert ist, gibt es Stufen, auf denen mehrere der 2^n Weltzustände auf z_i folgen können. Wenn auf z_i jeder der 2^n Zustände folgen kann, gibt es in diesem Augenblick (auf der Stufe i) keinerlei Determination; wenn nur ein einziger folgen kann, liegt auf der Stufe i vollständige Determination vor. Man kann sich das Modell in Gestalt eines topologischen Baumes vorstellen. Während nur ein einziger Ast den bisherigen Weltverlauf wiedergibt, entsprechen alle übrigen Äste des Baumes einer nicht realisierten Möglichkeit.

Die Entwicklung der physikalischen Grundlagenforschung der letzten fünfzig Jahre – besonders der bis heute irreduzibel probabilistische Charakter mikrophysikalischer Vorgänge, wie sie sich in der Quantenmechanik darstellen – wird i. allg. als Beleg dafür interpretiert, daß sich die Welt trotz zuverlässiger deterministischer makrophysikalischer Theorien nicht deterministisch verändert und der Weltverlauf also einen probabilistischen Charakter hat. Einige Autoren haben sogar versucht, daraus Argumente abzuleiten, die die Theorie menschlicher Willensfreiheit stützen sollen[30].

Zur schärferen Eingrenzung des Konsequentialismus ist es jedoch sinnvoll, zunächst einige grobe Vereinfachungen vorzunehmen: Wir nehmen an, daß der Weltverlauf ohne den Einfluß menschlichen Handelns streng deterministisch sei und menschliches Handeln nicht unter die den natürlichen Weltverlauf bestimmenden deterministischen Gesetze[31] falle. Menschliches Handeln wäre dann *in einem metaphysischen Sinne* frei, und die Verzweigungen des Baumes ergäben sich in diesem Fall ausschließlich aus dem Einfluß menschlichen Handelns auf den Weltverlauf. In der Regel hat eine Handlung (bzw. der mit der Handlung korrelierte Vorgang) eine gewisse zeitliche Ausdehnung. Idealisierend nehmen wir jedoch an, daß die betrachtete Handlung *punktuell* ist: Die Handlungsoptionen

[30] Vgl. dazu etwa P. Jordan, *Wie frei sind wir?*, Osnabrück 1971.

[31] Dieser Standpunkt wurde in jüngerer Zeit mit besonderer Radikalität von A. I. Melden, *Free Action*, London 1961 vertreten. Kant nimmt dagegen mit seiner Unterscheidung des phänomenalen und des noumenalen Ich einen differenzierteren Standpunkt ein: Einerseits kann sich der Mensch zur Verstandeswelt gehörend und damit unter dem Gesetz der Freiheit stehend betrachten, andererseits ist er Teil der Sinnenwelt und untersteht damit dem Kausalgesetz der Natur. Vgl. *Grundlegung zur Metaphysik der Sitten*, AA 450f.

legen, zum Zeitpunkt i vollzogen, einen bestimmten Nachfolgezustand z_{i+1} fest. Eine strategisch abwägende rationale Person berücksichtigt, welche Verzweigungen ihr bei der Wahl einer Handlungsoption in späteren Situationen zur Verfügung stehen. Im Idealfall, d. h. wenn ihr alle notwendigen Informationen zur Verfügung stehen und sie diese auch vollständig verwertet, wird sie sich nicht punktuell für die eine oder andere situationsbezogene Handlung, sondern für eine Gesamtstrategie entscheiden, die hinsichtlich des für sie relevanten Geschichtsabschnittes des Weltverlaufes optimal ist.

Die rationale einsame Einzelperson mit vollständigem Wissen und zeitlich festen subjektiven Wertungen hat, so scheint es, in einer so beschaffenen Welt keinerlei Entscheidungsprobleme. Sie wählt diejenige Folge von Handlungen, die angesichts des bestehenden Weltzustandes und der wirkenden deterministischen Gesetze einen Weltverlauf (eindeutig) festlegt, dessen Zeitintegral des subjektiven Wertes maximal ist. Sie wählt in dieser fiktiven Situation die Gesamtstrategie mit den besten Folgen: Sie handelt *im konsequentialistischen Sinne rational.*

Soweit man diese fiktive Situation in den Kontext der anthropozentrischen Hauptströmung abendländischer Ethiktradition stellt, scheint einem konsequentialistischen Modell von Handlungsrationalität nichts im Wege zu stehen: Da es nur die eine handelnde Person gibt, spielen die subjektiven Werte anderer Personen, die den Handelnden zu einer Abwägung eigener und fremder Interessen veranlassen und eine Modifikation der subjektiven Wertfunktion nach sich ziehen könnten, ebensowenig eine Rolle wie Aspekte der Gerechtigkeit und der Koordination. Hier scheint das Modell konsequentialistischer Handlungsrationalität eindeutig und überzeugend zu sein. Es liegt nahe, es auch auf komplexere Situationen zu übertragen.

2. Kapitel: Intentionen

Um das konsequentialistische Grundmodell weiter zu präzisieren, ist es unumgänglich, die Rolle der Intentionalität für menschliches Handeln zu klären. Das ist die Aufgabe dieses Abschnittes. Er setzt sich mit der Einbettung von Handlungsabsichten in eine generelle Konzeption der Intentionalität auseinander und unterscheidet drei verschiedene Arten von Handlungsabsichten.

§ 5 *Intentionalität*

Intentionalität[1] ist diejenige Eigenschaft geistiger Zustände und Ereignisse, durch die diese auf Gegenstände oder Sachverhalte in der Welt gerichtet sind oder von ihnen handeln[2]. „Intentionalität" und „Intensionalität" stehen dabei in einer engen Beziehung. In „intensionalen Kontexten" können extensionsgleiche Prädikate nicht in unter allen Umständen wahrheitswerterhaltend substituiert werden. Wenn wir Intentionen beschreiben, dann sind wir i. allg. auf intensionale Kontexte angewiesen.[3] In enger Anlehnung an Brentano hat R. Chisholm[4] eine Konzeption von sprachlich charakterisierter Intentionalität entwickelt, die die ursprüngliche metapsychologische Fragestellung in eine Frage nach der logischen Form von Sätzen über psychische Phänomene überführt und auf diesem Wege die enge Beziehung von Intentionalität und Intensionalität deutlich macht.

Überzeugungen sind Überzeugungen, daß etwas der Fall ist, Befürchtungen sind Befürchtungen, daß etwas geschieht, Wünsche sind Wünsche, etwas zu tun, etwas zu haben, etwas zu erleben, Vermutungen sind Vermutungen, daß etwas der Fall sein wird oder der Fall war etc. Intentionalität ist ein typisches Merkmal vieler, aber nicht aller mentalen Zustände: Es ist zweckmäßig, von einem engen Intentionalitätsbegriff auszugehen, der

[1] Chisholm fragt sich, wie ist objektive Referenz – bzw. Intentionalität – möglich? Wie ist es möglich, daß man seine Gedanken auf eine Sache richtet? Er entwickelt zu diesem Zweck eine platonische Ontologie, die von der Existenz *ewiger* („eternal") bzw. *abstrakter* Objekte ausgeht: *Es gibt* Eigenschaften, Relationen und Zustände („states of affairs"). Es ist hier nicht notwendig zu klären, ob diese Vorgehensweise notwendig und angemessen ist. Im folgenden bleiben daher diese ontologischen Fragen ausgeklammert, was jedoch dem Leser die Bereitschaft abverlangt, in Zweifelsfällen die ontologisierende Redeweise (Eine Handlung ist..., es gibt Absichten, die ... etc.) in eine Sachverhaltsredeweise (es ist der Fall, daß die Person absichtlich handelt...) zu übersetzen. Vgl. R. M. Chisholm *The First Person: An Essay on Reference and Intentionality*, Minneapolis 1981, bes. Kap. 2 und 4.

[2] Vgl. A. M. Anscombe, „The Intentionality of Sensation: A Grammatical Feature", in: *Analytical Philosophy*, hg. von R. J. Butler (1965). Vgl. a. J. R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge 1983, Kap. 1.

[3] Dieser Zusammenhang rechtfertigt nicht die Vorstellung, Intensionalität sei eine besondere Erscheinungsform von Intentionalität, vgl. W. Kneale, „Intentionality and Intensionality", *Arist. Soc. Suppl.* 42 (1968) 73–90 sowie im gleichen Band A. N. Prior, „Intentionality and Intensionality", 91–106.

[4] Chisholm gebraucht in Anlehnung an G. Frege den Ausdruck ‚indirect reference' für referentielle Opakheit. Vgl. R. Chisholm, „Sentences about Believing", in: *Intentionality, Mind and Language*, hg. von A. Marras, Chicago 1972, S. 31–51.

nicht jedem Bewußtsein als „Bewußtsein von etwas" Intentionalität zuschreibt: Mir ist zwar die Freude bewußt, aber die Freude ist nicht Freude von etwas[5]. Mit diesem engen Intentionalitätsbegriff ist die Identifizierung von „Bewußtheit" und „Intentionalität" ausgeschlossen. So sind Wünsche zwar grundsätzlich intentional, aber nicht immer bewußt – umgekehrt kann mir meine (nicht-intentionale) Niedergeschlagenheit bewußt sein. Obwohl Intentionalität nichts genuin Sprachliches ist[6] – es ist sinnvoll zu sagen, die lauernde Katze vor dem Loch *glaube*[7], daß sich darin eine Maus verberge – besteht, wie Searle überzeugend gezeigt hat, eine interessante Parallelität zwischen der Repräsentanz von Sachverhalten durch intentionale Zustände und der Repräsentanz von Sachverhalten durch Sprechakte.[8]

Die Sprechakttheorie[9] unterscheidet zwischen dem propositionalen Gehalt und der illokutionären Rolle, in der dieser präsentiert wird. In analoger Weise kann man bei intentionalen Zuständen den propositionalen Gehalt und die *Einstellung*, in der dieser präsentiert wird, unterscheiden. Ein

[5] Vgl. dazu die Behandlung selbstreferenzieller Formen der Intentionalität von J. Searle, *Intentionality*, Cambridge 1983, Kap. 2 und 3.

[6] Eine wichtige Strömung der Sprachphilosophie leitet Sprache logisch von Intentionalität her, während eine weitverbreitete Interpretation der Spätphilosophie Wittgensteins von der umgekehrten logischen Beziehung zwischen Sprache und Intentionalität ausgeht. Grundlegend für diese Strömung ist der Artikel von H. P. Grice, „Meaning", *Phil.Review* 66 (1957) 377–388 und „Utterer's Meaning and Intentions", *Phil.Rev.* 78 (1969) 147–177 sowie die Beiträge von J. Bennett, „The Meaning-Nominalist Strategy", *Foundations of Language* 10 (1973) 141–168 und D. Lewis, „Languages and Language", in: *Language, Mind and Knowledge* hg. von K. Gunnerson, Minneapolis 1975, S. 3–35 sowie ders., *Convention: A Philosophical Study*, Cambridge 1969. Wichtige Beiträge dieser Strömung sind in deutscher Übersetzung in dem Sammelband *Handlung, Kommunikation, Bedeutung*, Frankfurt/M. 1979, hg. von G. Meggle enthalten.

[7] Erst dadurch, daß wir auch höheren Tieren Intentionalität zusprechen, werden Formulierungen sinnvoll wie, das unruhige Verhalten des Hundes sei *verständlich*, da er aufgrund des vertrauten Motorengeräusches die Rückkehr der Familie *erwartet*, u. ä., vgl. R. Spaemann, „Naturteleologie und Handlung", *Zeitschrift für philosophische Forschung* 32 (1978), 481–493, S. 481: „Der wissenschaftliche Versuch, das Laufen des Hundes zum Freßnapf ohne das Wort „Hunger" und sein Gebell bei der Heimkehr des Herrn ohne das Wort „Freude" zu interpretieren, hat immer etwas Phantastisches."

[8] J. Searle (1983).

[9] Die Sprechakttheorie wurde von J. L. Austin mit dem Ziel entwickelt, die Rede vom „Gebrauch" eines sprachlichen Ausdrucks zu klären. Dies geschieht, indem nach den Handlungen gefragt wird, die der Sprecher mit einer bestimmten Äußerung vollzieht. Franz, der zu einem Besucher aus der Stadt sagt „dort steht ein Stier", vollzieht zunächst den lokutionären Akt zu sagen, daß dort ein Stier steht. Zugleich aber warnt er mit dieser Äußerung den Besucher – diese Handlung nennt Austin den illokutionären Akt. Und möglicherweise schüchtert Franz mit dieser Äußerung den Besucher ein – das nennt Austin den perlokutionären Akt. Austins Ansatz wurde von John Searle wesentlich erweitert und findet unterdessen in vielfältiger Weise Anwendung nicht nur in der Sprachphilosophie, sondern auch in Linguistik, Soziologie und Psychologie, vgl. J. L. Austin, *How to Do Things with Words* (1962), dt. Bearbeitung von E. v. Savigny, *Zur Theorie der Sprechakte*, Stuttgart 1972; J. R. Searle, *Speech Acts*, Cambridge 1969; dt. *Sprechakte: Ein sprachphilosophischer Essay*, Frankfurt/M. 1974; D. Wunderlich, *Studien zur Sprechakttheorie*, Frankfurt/M. 1976. Eine gute kritische Einführung in die Sprechakttheorie bietet E. v. Savigny, *Die Philosophie der normalen Sprache*, Frankfurt/M. 1974, Kap. 3 und W. Stegmüller, *Hauptströmungen der Gegenwartsphilosophie*, Bd. II, Stuttgart 1979, Kap. 1, § 3.

Beispiel für einen Sprechakt: „Ich warne Dich vor diesem Ereignis": Der propositionale Gehalt dieses Sprechaktes ist, daß dieses Ereignis eintreten wird; der Sprechakt ist der des Warnens. Ein Beispiel für einen intentionalen Zustand: „Ich hoffe, daß morgen die Sonne scheint": Der propositionale Gehalt ist, daß morgen die Sonne scheint, die Einstellung ist die des Hoffens.

So wie es Sprechakte gibt, die keinen propositionalen Gehalt haben[10], so gibt es intentionale Zustände ohne propositionalen Gehalt. Alle intentionalen Zustände sind durch bestimmte Einstellungen charakterisiert, aber man kann nicht in jedem Falle von *propositionalen* Einstellungen sprechen, da nicht alle intentionalen Zustände einen (vollständigen) propositionalen Gehalt haben. Der propositionale Gehalt von „Hans liebt Kalbsschnitzel" ist möglicherweise „Hans ißt Kalbsschnitzel", aber auch „Hans liebt Helga" charakterisiert einen intentionalen Zustand, ohne daß sich in diesem Fall (zwanglos) ein propositionaler Gehalt angeben ließe.

Auf einige, nämlich die *assertiven* Sprechakte (Behaupten, Vermuten etc.) lassen sich Prädikate wie „wahr" oder „falsch', anwenden. In analoger Weise können wir von assertiven intentionalen Zuständen sagen, sie seien zutreffend oder wahr bzw. falsch: Es gibt z. B. zutreffende Vermutungen. Andere, die *direktiven* Sprechakte wie Bitten oder Befehle, können nicht wahr oder falsch sein, aber sie werden erfüllt, indem sich eine Veränderung gegenüber dem ergibt, was geschehen wäre[11], wenn der Adressat der Bitte oder des Befehls sich nicht entsprechend verhalten hätte. Ihnen entsprechen direktive intentionale Zustände, wie Hoffnungen, Wünsche, Absichten. So wie der singuläre Sprechakt des Befehlens nicht wahr oder falsch ist, sondern erfüllt wird oder nicht, so ist der singuläre intentionale direktive Zustand des Wünschens nicht wahr oder falsch, sondern er erfüllt sich oder nicht. Auch dem kommissiven Sprechakttypus (Versprechen u. a.) entsprechen intentionale Zustände, etwa der des Verpflichtetfühlens, der Loyalität etc. Die enge Verbindung kommissiver Sprechakte und kommissiver intentionaler Zustände äußert sich vielfach sprachlich in „x empfindet y"-Wendungen: „Ich fühle mich verpflichtet", „ich fühle mich aufgefordert" etc. Die Unterscheidung in assertive, direktive und kommissive intentionale Zustände ist – wie bei Sprechakten – keine vollständige Disjunktion: Während man „Lieben" (unter bestimmten Umständen) noch unter die direktiven intentionalen Zustände rechnen kann (eine Liebe kann unerfüllt bleiben), ist das bei „Wundern" oder „Freuen" nicht der Fall, obwohl beide Zustände sicher nicht assertorisch oder kommissiv sind.

[10] Vgl. dazu E. v. Savignys Kritik der Austinschen Unterscheidung von lokutionären und illokutionären Akten, *Die Philosophie der normalen Sprache*, Frankfurt/M. 1974, S. 158ff.

[11] Diese Formulierung weist auf das jedem Handeln innewohnende *kontrafaktische* Element, vgl. G. H. v. Wright „Handlungslogik – Ein Entwurf" (1967), in: ders., *Handlung, Norm und Intention*, Berlin/New York 1977, sowie „Elemente der Handlungslogik", in: *Handlungstheorien – interdisziplinär*, Bd. I, München 1977–81.

Die für uns interessanteste Verknüpfung von Sprechakten und intentionalen Zuständen bezieht sich ausschließlich auf Sprechakte und intentionale Zustände mit propositionalem Gehalt. Zahlreiche Sprechakte mit propositionalem Gehalt stehen unter einer *Aufrichtigkeitsbedingung* ohne notwendigerweise *fehlzuschlagen*, wenn diese nicht erfüllt ist: Ich kann etwas empfehlen, auch wenn ich von den Qualitäten des Empfohlenen nicht überzeugt bin, aber dieses Empfehlen wäre unaufrichtig. Ähnliches gilt für die *assertiven* Sprechakte des Behauptens, Leugnens, Beschreibens etc.; *direktive*, wie des Befehlens (es ist kein aufrichtiger Befehl, wenn ich nicht wünsche, daß der Adressat den Befehl ausführt); und *kommissive*, wie etwa des Versprechens (es ist kein aufrichtiges Versprechen, wenn ich nicht beabsichtige, es einzuhalten). Ist die Aufrichtigkeitsbedingung erfüllt, so läßt sich von einem singulären Sprechakt auf einen singulären intentionalen mentalen Zustand schließen[12]. Wenn die Empfehlung (Sprechakt), die mir mein Nachbar gegeben hat, aufrichtig war, dann hat mein Nachbar bestimmte Überzeugungen (intentionaler Zustand) von den Qualitäten des Empfohlenen. Wenn sein Versprechen (Sprechakt) aufrichtig war, dann beabsichtigt er (intentionaler Zustand), das (oder bei mehreren Alternativen: etwas) zu tun, was das Versprechen erfüllt.

Die Aufrichtigkeitsbedingung ist aber nicht nur eine kontingente Bedingung, die – soweit erfüllt – Sprechakt und intentionalen Zustand verknüpft, vielmehr besteht zwischen Sprechakt und intentionalem Zustand eine besondere Beziehung, die man als eine Art „Repräsentation" auffassen kann. Das läßt sich daraus ersehen, daß Äußerungen, die einen Sprechakt mit der Negation des entsprechenden intentionalen Zustandes verknüpfen, widersprüchlich oder vielleicht angemessener: *selbstaufhebend* sind.

Beispiele:

a) assertiv: „Ich behaupte, Columbus war Spanier, aber ich bin nicht sicher, ob Columbus Spanier war";

b) kommissiv: „Ich verspreche, morgen zu kommen, aber ich habe nicht vor, morgen zu kommen";

c) direktiv: „Ich befehle Dir zu gehen, aber ich wünsche nicht, daß Du gehst".

In allen drei Fällen folgt der intentionale Zustand nicht logisch aus dem entsprechenden Sprechakt in dem Sinne, daß

[12] J. R. Searle, „Eine Taxonomie illokutionärer Akte", in: ders., *Ausdruck und Bedeutung*, Frankfurt/M. 1982 unterscheidet bei Sprechakten mit propositionalem Gehalt zwischen zwei Typen: „representatives" und „directives". „Representatives" führen Wahrhaftigkeitsbedingungen mit sich, während „directives" typischerweise Versuche des Adressaten nach sich ziehen, die tatsächliche Welt so zu beeinflussen, daß die entsprechende Proposition erfüllt ist. Vgl. dazu auch die Theorie J. Bennetts, über die (im wesentlichen) nicht-sprachliche Basis von Glaubens- und Intentionsbegriffen: *Linguistic Behaviour*, Cambridge 1976.

a) aus „x behauptet y" folgt „x ist sich sicher, daß y";
b) aus „x verspricht y" folgt „x beabsichtigt y";
c) aus „x befiehlt y" folgt „x wünscht y".

Dennoch „zerstört" die Abrede des entsprechenden intentionalen Zustandes den Sprechakt: Er kann in dieser Kombination nicht vollzogen werden. Während die Unglücksfälle („infelicities") von der Art Gamma 1 im Sinne Austins das Zustandekommen des Sprechaktes nicht unmöglich machen[13], bewirkt (wie diese Beispiele zeigen) das *Kundtun* – ja, in den meisten Fällen schon die bloße Publizität der Unredlichkeit – tatsächlich ein Fehlschlagen des Sprechaktes. Ein Befehl, von dem der Adressat weiß, daß er nicht ernst gemeint ist, wird i. allg. nicht ausgeführt. In einem bestimmten Sinne kann man sogar sagen, der Befehl wurde nicht gegeben, da „der Adressat ja wußte, daß er nicht ernst gemeint war". Die Konvention des Befehlens verlangt jedoch, daß der Befehlende seine Ausführung wünscht. In einem anderen Sinne kann man sagen, der Befehl wurde sozusagen „formal" gegeben, was jedoch bei explizit performativen Äußerungen auch für andere Arten von Fehlschlägen typisch ist[14]. Wenn Hans in freudigem Tonfall und spöttischer Miene sagt: „Ich warne Dich, dort steht ein Stier", dann hat er den Besucher „formal", aber eben nicht „wirklich" gewarnt. Es wäre unangemessen, in diesem Fall zu behaupten, er habe den Besucher im illokutionären, aber nicht im perlokutionären Sinn gewarnt, denn der Besucher *konnte* die Warnung unter diesen Umständen nicht ernst nehmen – es war insofern auch keine Warnung im illokutionären Sinne. Dieses Merkmal der Alltagssprache deutet darauf hin, daß die Interpretation des intentionalen Zustandes als etwas, das von dem entsprechenden Sprechakt *repräsentiert* wird, zutreffend ist.

Um eine mögliche Konfusion auszuschließen, muß darauf hingewiesen werden, daß dieser Repräsentationsbegriff ein anderer ist, als der von Searle für die Beziehung Sprechakt – propositionaler Gehalt verwendete[15]. Für Searle besteht jeder Sprechakt in einer Repräsentation von i. allg. nicht-mentalen Sachverhalten und Gegenständen: Die *Feststellung*, daß es regnet, repräsentiert denselben Sachverhalt wie die *Hoffnung*, daß es regnet. Der Befehl an Hans, das Zimmer zu verlassen, repräsentiert eine bestimmte Handlung von Hans und nicht den intentionalen Zustand des Befehlenden. Dieser Repräsentationsbegriff ist jedoch schon durch den Begriff des „propositionalen Gehaltes" abgedeckt. Sprechakte repräsentieren darüber hinaus in der hier gewählten Bedeutung in einem recht unmittelbaren Sinne mentale Zustände, wie die obigen Beispiele gezeigt haben. Viele Sprechakte haben einen propositionalen Gehalt, manche sind auf Dinge, Ereignisse oder Sachverhalte gerichtet – aber die Repräs-

[13] Vgl. *Theorie der Sprechakte*, a.a.O., vierte Vorlesung.
[14] Vgl. J. L. Austins sechste Vorlesung, a.a.O.
[15] Vgl. J. Searle, *Intentionality*, Cambridge 1983, Kap. 1, II.

sentation von etwas Nicht-Sprachlichem durch Sprachliches erfolgt auch in der Beziehung Sprechakt/mentaler Zustand. Diese Beziehung scheint universell zu sein: Jeder Sprechakt repräsentiert einen mentalen Zustand, in vielen Fällen (und das sind diejenigen, die uns hier interessieren) einen *intentionalen mentalen Zustand.*

Eine Warnung oder ein Befehl kann befolgt werden, ein Versprechen unerfüllt bleiben, eine Behauptung zutreffen: Nicht alle, aber zahlreiche Sprechakte haben *Erfüllungsbedingungen.* Wenn ein Sprechakt Erfüllungsbedingungen hat, dann hat der im Sprechakt repräsentierte intentionale Zustand ebenfalls Erfüllungsbedingungen, und diese Erfüllungsbedingungen sind identisch. Eine Behauptung (singulärer Sprechakt) ist wahr genau dann, wenn die in der Behauptung repräsentierte Überzeugung (singulärer intentionaler Zustand) zutrifft. Ein Versprechen (Sprechakt) ist genau dann erfüllt, wenn das im Versprechen repräsentierte Vorhaben (intentionaler Zustand) ausgeführt ist. Der Befehl ist genau dann befolgt, wenn der im Befehl repräsentierte Wunsch in Erfüllung gegangen ist. Handlungsabsichten charakterisieren eine spezielle Art von Intentionalität. Eine Handlungsabsicht ist ein bestimmter intentionaler mentaler Zustand, der in einer engen Beziehung zu Handlungen steht. Im allgemeinen ist eine Handlung die Erfüllungsbedingung einer Absicht. Man muß jedoch verschiedene Arten der Absichtlichkeit unterscheiden. Nur für einen Typus von Absichten gilt, daß eine Handlung die Erfüllungsbedingung der betreffenden Absicht ist.

§ 6 *Handlungen als intentionales Verhalten*

Die Absichtlichkeit einer Handlung ist Bestandteil des Handlungsbegriffes. Eine ‚unabsichtliche Handlung' ist ein Widerspruch in sich. Auch bei unabsichtlichem Verhalten kann man zwar auf die Frage, „Was hast du da getan?" z. B. antworten: „Oh, ich habe die Vase heruntergeworfen", aber wenn ich mich frage, was ich falsch gemacht habe, welche *Handlungen* der letzten halben Stunde falsch waren, werde ich nicht zu dem Ergebnis kommen, daß das Herunterwerfen der Vase eine falsche Handlung war, auch wenn ich diesen Vorgang bedauere (möglicherweise war es eine falsche Handlung, trotz des Getümmels so nah an diese Vase herangegangen zu sein). Wenn ich die Vase nicht unabsichtlich heruntergeworfen habe, dann habe ich sie *absichtlich*$_0$ heruntergeworfen – nur in diesem Fall entspricht diesem Vorgang eine (vermutlich tadelnswerte) Handlung.

Ein weiteres Beispiel: Jemand fährt betrunken Auto und überfährt einen Passanten, der an den Folgen seiner Verletzungen stirbt. Was hat er getan? Er hat einen Menschen getötet. Diese Tötung war sicher unabsichtlich. War sie eine Handlung? Auch hier ist auffallend, daß der Fahrer bei der Überlegung, welche seiner Handlungen der letzten halben Stunde falsch waren, nicht zu dem Ergebnis kommen wird, daß es falsch war, den

Passanten zu töten. Vermutlich hat er in der letzten halben Stunde viel falsch gemacht: Vielleicht hat er sich während der Fahrt intensiv mit seinem Mitfahrer unterhalten oder zuvor zwei Gläser über den Durst getrunken. Die Tötung selbst aber scheidet als eine falsche Handlung aus. Die Tötung des Passanten war keine Handlung des Fahrers. Offensichtlich sind es besonders die Handlungsbestandteile des Verhaltens, die einer Person *vorgehalten* werden können: Dem Fahrer werden bestimmte Aspekte seines Verhaltens zum Vorwurf gemacht nur soweit diese Aspekte mit Handlungen des Fahrers korrespondieren.

Dieser Sachverhalt bildet den Ausgangspunkt für H. L. A. Harts Vorschlag, den Handlungsbegriff eng mit dem der Verantwortung zu verknüpfen[16]. Dieser Vorschlag, der von der Rechtspraxis gestützt und mit einer rechtsphilosophischen Intention entwickelt wurde, ist jedoch zur Eingrenzung des Handlungsbegriffes keine Hilfe, da der Begriff der Verantwortlichkeit zu unscharf ist. So ist der Fahrer zwar in einem bestimmten Sinne verantwortlich für den Tod des Passanten, er hat ein entsprechendes Gefühl der Schuld, seine Versicherung wird für die Hinterbliebenenversorgung herangezogen etc., aber er wird andererseits nicht wegen Mordes oder Todschlags verurteilt, sondern z. B. wegen fahrlässiger Tötung (Fahrlässigkeit mit Todesfolge). Nur wenn Handlungen im Spiel sind, gibt es persönliche Verantwortlichkeit, aber die Relation „x ist verantwortlich für y" erlaubt für y auch Einsetzungen, die keine Handlungen von x sind, dazu zählen u. a. bestimmte (sicherlich nicht alle) Folgen einer Handlung von x.

Ein letztes Beispiel: Wenn ein geübter Autofahrer den Gang wechselt, tut er dies i. allg. nicht unabsichtlich – unabsichtlich geschieht es hingegen, wenn er zum Handschuhfach greifen will und dabei den Schaltknüppel berührt. Trotzdem muß der Handlung des Schaltens keine Absicht *vorausgehen* (bei einer bestimmten Tonhöhe des Motorgeräusches wechselt er normalerweise den Gang). Er hätte das Schalten auch unterlassen können, dennoch kann man nicht in der gleichen Weise sagen, er hätte, als er zum Handschuhfach greifen wollte, das Schalten auch unterlassen können. Insoweit ist das Schalten im ersten Fall ein kontrollierter, bewußter und absichtlicher$_0$ Vorgang, während im anderen Fall das Schalten ein unkontrollierter, unbewußter und unbeabsichtigter Vorgang ist. Diesen Unterschied kann man auch dadurch zum Ausdruck bringen, daß man im einen Fall sagt, „er schaltete mit Absicht", und im anderen „er schaltete ohne Absicht".

[16] H. L. A. Hart, „The Ascription of Responsibility and Rights", *Proc.Arist.Soc.* 44 (1948/49), abgedruckt in: *Logic and Language*, hg. von A. G. N. Flew, Oxford 1951, S. 145–166; vgl. dazu G. Pitcher, „Hart on Action and Responsibility", *Phil.Rev.* 69 (1960) 226–235, sowie J. Feinberg, „Action and Responsibility", in: ders., *Doing and Deserving*, Princeton/N.J. 1970, S. 119–151, umgearbeitet in: M. Black (Hg.), *Philosophy in America*, London 1965, S. 134–160.

Man könnte versucht sein, die mit den Handlungen verknüpften Vorgänge oder Resultate als Erfüllungsbedingungen der Absichten$_o$ zu interpretieren. Das scheitert jedoch an Beispielen der folgenden Art: Angenommen, die Erfüllungsbedingung der Absicht$_o$, die Bestandteil der Handlung „absichtlich schalten" ist, sei der Schaltvorgang als solcher. Dann wäre diese Absicht erfüllt, wenn etwa eine Vorrichtung zufällig im gleichen Zeitpunkt den Schaltvorgang bewirkt. Um paradoxen Konsequenzen dieser Art zu entgehen, bieten sich zwei Möglichkeiten an:

1) Als Erfüllungsbedingungen einer Absicht$_o$ werden komplexe Kausalvorgänge angenommen, so daß nicht nur der Schaltvorgang als solcher, sondern auch die kausalen Bedingungen seines Zustandekommens Bestandteil der Erfüllungsbedingungen der entsprechenden Absicht$_o$ sind.

2) Man akzeptiert die Absichten$_o$-Sprechweise nur als Variante der „absichtlich tun"-Sprechweise, wobei „i hat x absichtlich getan" nicht impliziert, daß i eine Absicht in der üblichen Bedeutung hatte; denn Absichten haben als intentionale Zustände (propositionale) Erfüllungsbedingungen, während das gleiche für Absichten$_o$ nicht gilt.

Der erste Lösungsversuch kann nicht gelingen, denn es scheint für jede Erfüllungsbedingung möglich zu sein, einen Vorgang zu konstruieren, der diese Erfüllungsbedingung *kausal* bewirkt, ohne daß diesem Vorgang eine absichtliche Handlung entspricht. Für die zweite Lösung spricht außerdem, daß man sprachlich für „i tat x absichtlich" zwar auch sagen kann „i tat x mit Absicht", aber die Frage „welche Absicht hatte er?" in diesen Fällen keine vernünftige Antwort hat bzw. sich vernünftige Antworten auf Absichten anderen Typs beziehen, nämlich entweder auf *vorausgehende* oder auf *motivierende* Absichten.

§ 7 *Drei Arten von Intentionen*

Etwas kann absichtlich$_o$ getan werden, ohne daß der Tat eine Absicht *vorausgeht*, wie das Schalt-Beispiel deutlich macht: Nicht jede Handlung hat eine *vorausgehende Absicht*. Handlungen mit vorausgehenden Absichten sind dadurch gekennzeichnet, daß der Handelnde vorher wußte, was er tun wird. Absichtliche$_o$ Handlungen ohne vorausgehende Absichten werden vollzogen, ohne daß der Handelnde zuvor wußte, daß er sie vollziehen wird.

Wenn Bill beabsichtigt, Carillo zu töten, wird diese Absicht erfüllt, wenn Bill Carillo tötet. Sie wird jedoch nicht erfüllt, wenn Carillo auf andere Weise, etwa durch Steinschlag zu Tode kommt, ja, selbst dann nicht,

wenn Bill Carillo unabsichtlich tötet[17]. Die Erfüllungsbedingung einer vorausgehenden Absicht ist nicht ein Vorgang oder ein Ereignis, sondern eine absichtliche$_o$ Handlung. Absichten dieses Typs sind kein „begrifflicher" Bestandteil der Handlung[18].

Formulierungen der Art „er hat h mit der Absicht x getan" verweisen auf einen dritten Typ von Absichten, die wir *„motivierende Absichten"* nennen. Diese motivieren die infrage stehende Handlung, obwohl die Handlung nicht Erfüllungsbedingung ist, und a fortiori ist eine motivierende Absicht nicht Bestandteil der Handlung. Manche motivierenden Absichten haben jedoch in vielen Fällen *andere* Handlungen als Erfüllungsbedingungen – in jedem Fall sind die Erfüllungsbedingungen motivierender Absichten einer Person kausal oder begrifflich eng mit den Handlungen dieser Person verbunden.

Wenn ich jemandem eine Medizin einflöße mit der motivierenden Absicht, daß er gesund wird, so ist die Erfüllungsbedingung dieser Absicht eine mögliche und erhoffte Konsequenz meiner Handlung, aber nicht die Handlung selbst. Sorgfältiger formuliert wird jedoch auch in diesem Beispiel die Handlungsnähe deutlich: Meine motivierende Absicht ist, zur Gesundung der Person durch diese Handlung *beizutragen*. Die Absicht hat also nicht die Proposition „daß er gesund ist (oder gesund wird)", sondern die Mitwirkung des Handelnden am Gesundungsprozeß (als Erfüllungsbedingung) zum Gegenstand.

Wir unterscheiden also:

1) Absichten$_o$ als Bestandteil der jeweiligen Handlungen. Absichten$_o$ nennen wir auch *‚handlungskonstituierende Absichten'*.

2) Absichten$_v$ als der Handlung *vorausgehende Absichten*. Diese sind nicht handlungskonstituierend: Die gleiche Handlung kann auch ohne diese Absicht vollzogen werden. Erfüllungsbedingung von vorausgehenden Absichten sind absichtliche$_o$ Handlungen.

[17] Vgl. R. M. Chisholm, „Freedom and Action", in: *Freedom and Determinism*, hg. v. K. Lehrer, New York 1966.

[18] Absichten$_v$ kommen daher als kausale Determinaten von Handlungen in Betracht, auf diesen Typus von Absichten läßt sich der gegen den handlungstheoretischen Kausalismus vorgebrachte Einwand – Absichten scheiden als Ursachen von Handlungen aus, da Absichten mit Handlungen logisch verknüpft sind – nicht anwenden. Sehr pointiert wird dieses Argument von A. I. Melden, *Free Action*, London 1961, bes. Kap. 11 und 12 entwickelt. Andere wichtige Kritiker des Kausalismus, die sich ebenfalls in erster Linie auf das Argument der sog. logischen Beziehung stützen, sind: G. E. M. Anscombe, *Intention*, Oxford 1957; P. Foot, „Free Will as Involving Determinism", in: *Free Will*, hg. von S. Morgenbesse u. J. J. Walsh, Englewood Cliffs 1962; A. Kenny, *Action, Emotion and Will*, London 1963; C. Taylor, *The Explanation of Behavior*, London 1964; D. Bennett, „Action, Reason, and Purpose, *J. Phil.* 62 (1965) 85–96; R. Taylor, *Action and Purpose*, Englewood Cliffs 1966; F. Stoutland, „The Logical Connection Thesis", *Am.Phil.Quart. Monograph* 4 (1970). Eine systematische kritische Diskussion antikausalistischer Argumente enthält A. Beckermann, *Gründe und Ursachen*, Kronberg/Ts. 1977, Interessante Argumente für eine kausalistische Auffassung gibt: W. Gean, „Reasons and Causes", *Rev. of Metaphysics* 19 (1965/66) 667–688.

3) *Motivierende Absichten*$_m$, die weder konstituierender Bestandteil von Handlungen sind noch Handlungen als Erfüllungsbedingungen haben.

Eine vorausgehende Absicht$_v$ motiviert nicht dazu, eine Handlung zu vollziehen, die Erfüllungsbedingung dieser vorausgehenden Absicht ist. Sie kann zu anderen Handlungen motivieren, etwa zu solchen, deren wahrscheinliche Konsequenz eine Handlung ist, die Erfüllungsbedingung dieser Absicht$_v$ ist. Eine vorausgehende Absicht hat eher den Charakter einer Entscheidung. „Ich habe die Absicht, x zu tun", besagt – je nach Situation – in etwa das gleiche wie „Ich habe mich entschieden, x zu tun, soweit ich nicht von irgendwelchen mir jetzt noch unbekannten Umständen daran gehindert werde". Vorausgehende Absichten können weder als Motivation noch als Grund für die entsprechende[19] Handlung angeführt werden, da die enge Verwandtschaft mit Entscheidungen es mit sich bringt, daß sich die Begründungsbedürftigkeit einer Handlung auf die durch sie erfüllten Absichten$_v$ überträgt.

Absichten$_m$ dagegen können jederzeit als Gründe für Handlungen (und ihre Absichten$_v$) angeführt werden. Absichten$_m$ bezüglich einer Handlung H sind dabei häufig Absichten$_v$ bezüglich einer anderen Handlung H′, wobei H′ Konsequenz (nicht Resultat) einer Handlung H ist. Diese Kette führt mit logischer Notwendigkeit zu einer Abschlußhandlung H_z, als deren Grund nur solche Absichten$_m$ infrage kommen, die bezüglich keiner Handlung vorausgehende Absichten sind[20].

Die eingeführten begrifflichen Unterscheidungen lassen sich an einem Beispiel illustrieren: Eine Krankenschwester (S) geht zum Medizinschrank, nimmt ein Fläschchen heraus, geht zurück zum Patienten (P), flößt ihm die Flüssigkeit ein, setzt sich neben das Bett, wartet eine Weile, erhebt sich dann, geht zum Waschbecken, bereitet kalte Wadenwickel vor etc. Mit etwas Phantasie kann man sich eine lange Reihe von Vorgängen, Aktivitäten und Handlungen vorstellen, die sich über mehrere Tage erstrecken: „Die Schwester pflegte den Patienten". Diese tagelang andauernde Aktivität beinhaltet eine Reihe absichtlicher$_o$ Handlungen. Daneben tut S immer einmal wieder etwas unabsichtlich$_o$ – es handelt sich dann

[19] „Entsprechende" ersetzt den umständlichen Ausdruck „für eine Handlung, die Erfüllungsbedingung der vorausgehenden Absicht ist", bzw. korrekter: „für diejenige Handlung, die infrage steht und die eine derjenigen Handlungen ist, die Erfüllungsbedingung der vorausgehenden Absicht sind".

[20] Der Grund einer Abschlußhandlung muß natürlich keine Absicht$_m$ sein, auch andere intentionale Zustände, auch solche ohne propositionalen Gehalt, können der Grund einer Handlung sein. Wir konzentrieren uns dennoch zunächst auf die Rolle von Absichten, da dieser besondere Typ intentionaler Zustände nach konsequentialistischer Auffassung in besonders enger Weise mit Handlungsrationalität verknüpft ist: Es ist typisch für eine rationale Person, aufgrund von Absichten und nicht aufgrund bloßer momentaner Anwandlungen, Gefühlsaufwallungen etc. zu handeln. Die Tatsache, daß „Absichten" dem Akteur bewußt sind, ist offensichtlich sogar analytisch wahr, während die meisten anderen als Handlungsdeterminanten infrage kommenden intentionalen Einstellungen der handelnden Person nicht bewußt sein müssen.

nicht um Handlungen von S, sondern um bloßes Verhalten. Meist sind es Bestandteile absichtlicher$_o$ Handlungen, die – da häufig vollzogen – unabsichtlich, sozusagen automatisch ablaufen. Nur ein Teil der absichtlichen$_o$ Handlungen, die sich zusammen mit den unabsichtlichen$_o$ Aktivitäten zu einem komplexen Pflegevorgang zusammensetzen, sind Erfüllungsbedingungen vorausgehender Absichten. Wenn S den Löffel in einer bestimmten Weise bewegt, damit kein Tropfen verlorengeht, dann korrespondiert dem keine vorausgehende Absicht; denn S ist dieser Vorgang so vertraut, daß sie keine vorausgehende Absicht benötigt, um den Löffel in der geforderten Weise zu bewegen. Die Absicht$_v$ beschränkt sich darauf, P die Medizin einzuflößen, und ist durch die Handlung des Einflößens der Flüssigkeit erfüllt. Die besondere Ausführung dieser Handlung ist absichtlich$_o$, ihr geht aber keine Absicht$_v$ voraus. S ist aufgestanden, um zum Medizinschrank zu gehen. Dieses Aufstehen ist sicher eine Handlung. Sie geschah nicht unabsichtlich, insofern war es eine absichtliche$_o$ Handlung. Es kam zu dieser Handlung, weil S die Medizin holen wollte: S hatte die Absicht, die Medizin zu holen – S stand in der Absicht auf, die Medizin zu holen. S hatte vermutlich nicht erst die Absicht aufzustehen und erfüllte dann diese Absicht, indem sie aufstand. Vielmehr gab es für diese Handlung keinerlei vorausgehende Absicht. Die Absicht, die Medizin zu holen, ist daher für die Handlung des Aufstehens nur motivierende, nicht vorausgehende Absicht. Ihre Erfüllungsbedingung ist die relativ komplexe absichtliche Handlung des Medizinholens.

Über die Gesamtaktivität des Plfegens spannt sich ein feinmaschiges und mehrschichtiges Netz von vorausgehenden und motivierenden Absichten. Je geübter die Pflegerin ist, desto größer werden die Maschen des Netzes (ein immer größerer Anteil der Gesamtaktivität geschieht ohne vorausgehende Absichten, wenn auch i. allg. nach wie vor absichtlich$_o$), und das Netz kommt mit weniger Schichten aus. Komplexere Aktivitäten werden zu Basishandlungen, oder anders formuliert: Der Anteil derjenigen Handlungen, die nicht aus anderen Handlungen zusammengesetz sind, nimmt zu. Die Gesamtaktivität des Pflegens ist sicher absichtlich (S pflegt P nicht aus Versehen oder unabsichtlich), insofern kann man – auch wenn es etwas künstlich wirkt – diese Gesamtaktivität als eine komplexe Handlung ansehen. Diese Gesamtaktivität ist von der Absicht$_m$ geleitet, dazu beizutragen, daß P gesund wird. Diese Absicht strukturiert das ganze Netz der übrigen Absichten. Die übrigen Absichten haben gegenüber dieser Absicht Mittelcharakter, das gilt sowohl für vorausgehende wie für motivierende Absichten. Es gibt offensichtlich keine Handlung, die Erfüllungsbedingung dieser übergeordneten Absicht$_m$ ist, soweit man „Beizutragen-daß-er-gesund-wird" nicht selbst wiederum als eine Handlung auffassen möchte. Der übergeordneten Absicht korrespondiert der übergeordnete Wunsch, daß P gesund wird. Erst dieser korrespondierende Wunsch verankert das die Gesamtaktivität strukturierende Netz von Ab-

sichten endgültig mit der äußeren Welt; denn die Erfüllungsbedingung dieses intentionalen Zustandes ist eine natürliche Proposition, während die Erfüllungsbedingungen von Absichten niemals handlungsunabhängig sind. S kann *sich wünschen*, aber *nicht beabsichtigen*, daß P gesund wird. Wenn S weiß, daß sie persönlich keinen Beitrag zu P's Gesundung leisten kann, dann macht es auch keinen Sinn, wenn S sagt, sie habe die Absicht, P gesund zu machen. S kann jedoch den Wunsch haben, daß P gesund wird, auch wenn S keinen Beitrag dazu leisten kann. Aber wenn S den Wunsch hat, daß P gesund wird, und S die Möglichkeit hat, (handelnd) dazu beizutragen, und S dies auch weiß, dann hat S eine entsprechende übergeordnete (motivierende) Absicht, aus der heraus sich das ganze Netz motivierender und vorausgehender Absichten entwickelt, das ihr (Pflege-) Handeln leiten wird[21].

[21] Die auf Aristoteles (Nikomachische Ethik 1146ff.) zurückgehende und insbesondere von G. H. v. Wright fortentwickelte Idee eines praktischen Schlusses, in dem von einem Handlungswunsch (einer konativen Prämisse) und einer der handelnden Person bekannten empirischen Bedingung auf eine entsprechende Handlung dieser Person geschlossen wird, kann man als logische Verbindung von Absichten rekonstruieren – vgl. Verf., „Der praktische Schluß des Aristoteles und das revealed preference-Konzept der rationalen Entscheidungstheorie", in: *Praktische Logik*, hg. v. P. Klein, Göttingen 1990, S. 203–216.

3. Kapitel: Bayes'scher Konsequentialismus

Die konsequentialistische Interpretation von Handlungsrationalität hat durch die Entwicklung der Entscheidungs- und Spieltheorie seit Mitte dieses Jahrhunderts eine theoretische Grundlage erhalten, mit der sich jede Kritik des Konsequentialismus auseinandersetzen muß, will sie sich nicht dem Vorwurf aussetzen, nur für ältere und schwächere Formen des Konsequentialismus relevant zu sein.

§ 8 Das Bayes'sche Kriterium

Wir nennen eine Entscheidung *entscheidungstheoretisch rekonstruierbar* genau dann, wenn sie im Sinne des Bayes'schen Kriteriums rational ist. Das Bayes'sche Kriterium lautet: Diejenige Entscheidung ist rational, die den (subjektiven) Erwartungswert des (subjektiven) Nutzens des Handelnden maximiert.

Um dieses Kriterium beurteilen zu können, betrachten wir zwei Theorien der Entscheidungsrationalität: T1 und T2. Als heuristisches Mittel führen wir den Begriff der *„(idealen) rationalen Person bezüglich einer Theorie T"* ein: die Entscheidungen der (idealen) rationalen Person bez. T erfüllen die von T geforderten Bedingungen der Entscheidungsrationalität.

Die rationale Person bez. T1 hat in jeder Situation eine bestimmte Menge von Entscheidungsalternativen zur Auswahl (T1a). Sie hat darüber hinaus eine (wohldefinierte) Nutzenfunktion, d. h. sie kann jedem Ereignis (Zustand, Weltverlauf) eine reelle Zahl als (kardinales) Maß ihrer (subjektiven) Vorliegen und Wertungen zuordnen (T1b). Und sie hat zu jedem Zeitpunkt eine kohärente (subjektive) Wahrscheinlichkeitsverteilung über alle Mengen zukünftiger Ereignisse (Zustände, Weltverläufe) (T1c). Die rationale Person bez. T1 maximiert in jedem Zeitpunkt den subjektiven Erwartungswert ihrer subjektiven Bewertungsfunktion (Bayes-Kriterium).

T1 verlangt viel von einer rationalen Person. Da eine konsistente Bewertung von Ereignissen ihre kausalen Verknüpfungen zu berücksichtigen hat, muß die Person imstande sein, ganze Weltverläufe in ihrer Bewertung zu berücksichtigen, speziell muß sie die Konsequenzen ihrer Entscheidungen bis weit in die Zukunft hinein abschätzen können. Zumindest muß sie in der Lage sein, den mit einer Handlung verbundenen denkbaren Weltverläufen konsistente Wahrscheinlichkeiten zuzuordnen: Über die Mengen beliebiger logisch möglicher Weltverläufe der Zukunft ordnet T1 einer rationalen Person eine kohärente Wahrscheinlichkeitsverteilung zu. Darüber hinaus muß die bez. T1 rationale Person die unterschiedlichen Kriterien und Aspekte, die ihre Bewertung prägen, in einer einzigen numerischen Zuordnung zusammenfassen, was die Abwägung schwer vergleichbarer Größen – wie z. B. die zwischen eigenem Wohlergehen und dem Maß der Gerechtigkeit einer Verteilung – einschließen kann.

Es liegt daher auf der Hand, daß die Anwendung von T1 sowohl als handlungsleitende normative Theorie, wie als analytisches Instrument zur „rationalen Rekonstruktion" menschlichen Entscheidungsverhaltens Schwierigkeiten macht[1]. Möglicherweise ist die ideale rationale Person bez. T1 eine fiktive Gestalt, die als Orientierung für reales Handeln nicht geeignet ist. Ebenso könnte der Versuch, Entscheidungen realer Personen im Hinblick auf T1 zu rekonstruieren fehlschlagen, da die Ähnlichkeiten zwischen realen Personen und der idealen rationalen Person bez. T1 zu geringfügig sind. Aber selbst wenn es sich so verhält, ist es nicht ausgeschlossen, daß T1 in idealtypischer Weise Entscheidungsrationalität charakterisiert. Es gibt in der Philosophie unterschiedliche Auffassungen darüber, ob ein derart abstraktes Modell noch sinnvoll sein kann. Die Erfolge der theoretischen Naturwissenschaften sprechen jedoch für eine große methodische Toleranz gegenüber typisierender Abstraktion. Mit der Kritik an der mangelnden Anwendbarkeit des Bayes'schen Modells ist noch nicht entschieden, daß der Bayes'sche Konsequentialismus keine adäquate idealtypische Charakterisierung von Entscheidungsrationalität ist.

Betrachten wir eine zweite Theorie der Entscheidungsrationalität T2. Diese Theorie stellt nicht ein *Kriterium* für die Auswahl der rationalen Entscheidung auf, sondern formuliert folgende *Mindestbedingungen* an die Präferenzen einer rationalen Person.

Seit X eine Menge von Zuständen, auf die sich die Präferenzen der rationalen Person bez. T2 beziehen. Wenn die Person eine Präferenz für die Alternative x gegenüber der Alternative y hat, schreiben wir abgekürzt: $x > y$ („strikte Präferenz"). Von einer „schwachen Präferenz" für x gegenüber y sprechen wir, wenn die Person entweder x präferiert oder zumindest zwischen beiden Alternativen x und y indifferent ist: $x \geq y$. Wenn man die strikte Präferenz als Grundbegriff wählt, läßt sich die schwache Präferenz definieren: $x \geq y \equiv \neg\, x < y$. Allerdings ist diese Definition nur sinnvoll, wenn die Präferenzrelation $\geq$ vollständig ist, daher ist es günstiger als Grundbegriff die schwache Präferenz zu wählen und die starke Präferenz per Definition einzuführen: $x > y \equiv x \geq y \wedge \neg\, y \geq x$.

Die schwachen Präferenzen einer rationalen Person sollen nun folgende Eigenschaften haben (bzw. Bedingungen erfüllen):

1) Reflexivität: $\forall\, x \in X : x \geq x$.

2) Vollständigkeit (Konnexität): Es wird angenommen, daß die Person für beliebige Alternativen x, y aus X eine Präferenz hat: Entweder zieht die Person x y gegenüber vor, oder sie zieht die Alternative y gegenüber x vor, oder sie ist zwischen beiden indifferent. $\forall\, x, y \in X : [x \geq y \vee y \geq x]$.

[1] H. Simon geht so weit, festzustellen „Conceptually, the SEU (=subjective expexted utility) model is a beautiful object deserving a prominent place in Plato's heaven of ideas. But vast difficulties make it impossible to employ it in any literal way in making actual human decisions", *Reason in Human Affairs*, Oxford 1983.

3) Transitivität: Die dritte Bedingung verlangt von der Person P, daß sie eine Präferenz für x gegenüber z hat, wenn sie x y gegenüber vorzieht und y z gegenüber vorzieht: $\forall x, y, z \in X : [x \geq y \wedge y \geq z \rightarrow x \geq z]$.

Diese drei Bedingungen können dadurch verschärft werden, daß man der Person nicht nur Präferenzen über die Zustände aus X, sondern auch Präferenzen über *Wahrscheinlichkeitsverteilungen* über X unterstellt. Diese Verschärfung scheint insofern zulässig zu sein, als man im Alltagsleben vielfach nicht zwischen sicheren Alternativen entscheidet, sondern zwischen Alternativen, deren Folgen nur (in unterschiedlichem Maße) wahrscheinlich sind. Im obigen Beispiel: Wenn ich von München nach Frankfurt will und das Flugzeug nehme, dann muß ich mit einer bestimmten Wahrscheinlichkeit mit einem schweren Unwetter oder gar einer vereisten Piste in Frankfurt rechnen, die mich mehr Zeit kostet, als wenn ich mit der Bahn gefahren wäre. Dennoch wird in vielen Fällen eine Reise mit dem Flugzeug vorgezogen, was damit zusammenhängt, daß eine Abwägung vorgenommen wird, die die jeweiligen Konsequenzen der betreffenden Handlung bei gegebenen Umständen mit den Wahrscheinlichkeiten gewichtet, die diesen Umständen beigemessen werden.

Wir nennen eine Wahrscheinlichkeitsverteilung über X eine „Aussicht" a. wenn gilt, daß die Summe der Wahrscheinlichkeiten der Zustände 1 ergibt und sich die Zustände gegenseitig ausschließen – $p\,(x \& y) = 0$ für $x \neq y$ (einfachheitshalber gehen wir davon aus, daß die Alternativenmenge endlich ist, wodurch die Phantasie nicht wesentlich eingeschränkt wird, da es sehr große natürliche Zahlen gibt).

Sei A die Menge aller Aussichten über X. Nun wird nicht nur gefordert, daß die Präferenzen der Person in X vollständig, reflexiv und transitiv sind, sondern daß die gleichen Eigenschaften auch für die Präferenzen bezüglich A gelten. Wenn man einen Zustand x aus X mit der Aussicht identifiziert, die jeder anderen Alternative aus X die Wahrscheinlichkeit 0 und der Alternative x die Wahrscheinlichkeit 1 zuordnet, dann sind in dieser Forderung die drei erstgenannten Forderungen logisch enthalten. Weitere vier Bedingungen vervollständigen die Charakterisierung rationaler Präferenzen:

1) Es muß gefordert werden, daß die Person zwischen zwei Aussichten indifferent ist, wenn sich eine Aussicht in die andere durch mathematische Umformung nach dem Wahrscheinlichkeitskalkül überführen läßt (Reduktion).

2) Weiterhin soll gelten, daß die Person, wenn sie zwischen einem x aus X und einer Aussicht a indifferent ist, überall x durch a ersetzen kann, ohne daß sich an ihren jeweiligen Präferenzen etwas ändert (Substituierbarkeit).

3) Sei x_a der für die Person beste und x_z der für die Person schlechteste Zustand – $\forall x \in X : x_a > x > x_z$; $x_a, x_z \in X$ – dann muß es für jeden Zustand

x aus X eine Aussicht a geben, die nur x_a und x_z Wahrscheinlichkeiten größer O zuordnet, so daß die Person zwischen x und a indifferent ist (Stetigkeit). Auch diese Bedingung ist durchaus plausibel; denn wenn die Person x_a gegenüber x vorzieht, dann muß es eine, möglicherweise sehr hohe Wahrscheinlichkeit für x_a in a geben, bei der die Person a gegenüber x vorzieht. Ebenso müßte es umgekehrt eine entsprechend hohe Wahrscheinlichkeit für x_z geben, bei der die Person x gegenüber a vorzieht. In diesem Kontinuum sich ändernder Wahrscheinlichkeiten zwischen x_a und x_z muß es eine Verteilung geben, bei der die Person zwischen a und x indifferent ist (Stetigkeit).

$\forall x \in X : \exists p,\ 0 \leq p \leq 1 : [\langle p/x_a \ \& \ (1-p)/x_z\rangle = x]$,

„=" steht hier für Indifferenz, spitze Klammern symbolisieren Aussichten (hier mit der Wahrscheinlichkeit p für x_a und der Wahrscheinlichkeit $(1-p)$ für x_z).

4) Wenn die Person mit zwei Aussichten konfrontiert wird, deren Wahrscheinlichkeiten sich auf ein gleiches Paar von Zuständen aus X beziehen, dann sollte sie diejenige Aussicht vorziehen, die für den präferierten Zustand eine höhere Wahrscheinlichkeit beinhaltet:

$\forall x, y \in X : [x > y \rightarrow (p \geq p' \rightarrow \langle p/x \ \& \ (1-p)/y\rangle \geq \langle p'/x \ \& \ (1-p')/y\rangle)]$ (Monotonie).

Die Theorie T2 besagt, daß die Präferenzen einer rationalen Person diese oben aufgeführten Bedingungen erfüllen.

Das zentrale Theorem der ökonomischen Nutzentheorie beinhaltet, daß die beiden Theorien T1 und T2 *logisch äquivalent* seien. Wenn dem so ist, wären wir mit einem starken Argument für den Konsequentialismus als adäquate Theorie der Entscheidungsrationalität konfrontiert; denn die Zurückweisung der konsequentialistischen Theorie T1 würde die Zurückweisung mindestens eines der Postulate von T2 erfordern: Der Bayes'sche Konsequentialismus würde in der Gestalt der Postulate von T2 seine Rechtfertigung erfahren. Bevor wir uns mit dieser Behauptung auseinandersetzen, ist das Theorem darzustellen, welches beansprucht, die Äquivalenz von T1 und T2 aufzuzeigen.

Das Theorem (Nutzentheorem) besagt folgendes: Genügen die Präferenzen einer Person den Bedingungen von T2, dann gibt es eine reellwertige Funktion u über A, so daß gilt:

1) $\forall a, a' \in A : [a \geq a' \leftrightarrow u(a) \geq u(a')]$.

2) Eine solche Funktion u aus (1) ist linear, d.h. $\forall a, a' \in A : \exists p,\ 0 \leq p \leq 1 :$ $[u(\langle p/a \ \& \ (1-p)/a'\rangle) = pu(a) + (1-p)\, u(a')]$.

3) eine solche Funktion ist durch die Präferenzen der betreffenden Person bis auf lineare Transformation eindeutig bestimmt. D.h.: Sind u und u' zwei Funktionen, die (1) erfüllen, dann gilt:

$\exists r, s \in \mathbb{R} : \forall a \in A : [u'(a) = r\, u(a) + s]$.

In der Literatur wird eine Funktion u dieser Art meist als „Nutzenfunktion" bezeichnet. Ob dies eine sinnvolle oder eine irreführende Benennung ist, hängt vom „Charakter" der dieser Funktion zugrundeliegenden Präferenzen ab. Wenn sie die subjektiven Interessen wiedergeben oder wenigstens eigenorientiert sind, erscheint diese Bezeichnung als angemessen, wenn sie dagegen moralischen Überzeugungen Ausdruck geben, ist diese Bezeichnung zumindest irreführend. Der Sachverhalt, daß Präferenzen durch eine Funktion u mit den o. g. Eigenschaften repräsentiert werden können, sagt nur etwas über ihre Kohärenz angesichts unsicherer Alternativen, aber nichts über ihre Motive (ihre inhaltliche Bestimmung) aus, denn es spricht nichts gegen die Vermutung, daß auch moralisch motivierte Präferenzen einer rationalen Person den Bedingungen von T2 genügen.

Das Theorem hat den Charakter eines *Metrisierungs-Theorems:* Es stellt fest, daß sich aufgrund eines komparativen Begriffes der schwachen Präferenz unter bestimmten Bedingungen eine quantitative Bewertungsfunktion (bis auf lineare Transformation) eindeutig bestimmen läßt. Jeder, der die genannten Bedingungen von T2 für Mindestbedingungen idealen rationalen Verhaltens hält, ist aus logischen Gründen gezwungen zu akzeptieren, daß eine rationale Person in ihrem Entscheidungsverhalten die Bewertungsfunktion u maximiert[2].

Wenn die Funktion u des Theorems mit der subjektiven Bewertungsfunktion von T1b identisch wäre und die Menge der subjektiven Wahrscheinlichkeitsverteilungen von T1c entspräche, dann würden die Rationalitätstheorien T1 und T2 logisch äquivalent. Äquivalenz von T1 und T2 hieße jedoch auch: Jeder, der die Postulate von T2 für vernünftige Bedingungen rationalen Verhaltens hält, muß den Konsequentialismus in der Fassung T1, „die Bayes'sche Fassung des Konsequentialismus", akzeptieren.

Um beurteilen zu können, ob T1 und T2 tatsächlich äquivalent sind, ist es erforderlich, die Interpretation der verwendeten Grundbegriffe zu prüfen. In T1 war die Rede von einer der rationalen Person zugeordneten subjektiven Bewertung von Ereignissen. Es wurde in T1 nicht gesagt, wie man die Funktion bestimmen kann, die dieser Bewertung entspricht. Das Nutzentheorem wird i. allg. als eine solche Möglichkeit interpretiert. Das würde allerdings die extensionale Äquivalenz der beiden Theorien bereits voraussetzen und ist uns daher keine Hilfe.

[2] Ein Beweis, der sich der hier eingeführten Terminologie bedient, findet sich in L. Kern/J. Nida-Rümelin: *Logik kollektiver Entscheidungen*. München, 1994 i. E., Kap. 2.

§ 9 Konsequentialistische Handlungsgründe

Angenommen, die Aussage ist wahr: „Die Person P hat Grund für h", bzw. genauer „P hat zum Zeitpunkt t Grund, die Handlung h zu vollziehen": G(P,t,h).

1) Folgt aus G(P,t,h), daß es etwas gibt (ein Ziel Z), das P durch h erreichen möchte?

2) Folgt aus G(P,t,h), daß P der Grund G bekannt ist?

3) Folgt aus G(P,t,h), daß es keine andere Handlung h' geben kann, für die es ebenfalls einen Grund für P gibt: G(P,t,h')?

Es ist das zentrale Ziel dieser Arbeit aufzuzeigen, daß die erste Frage zu verneinen ist. Die Begründung dafür ergibt sich jedoch erst vor dem Hintergrund der ethischen Kritik des Konsequentialismus in Kap. III und den handlungstheoretischen Überlegungen in Teil IV.

Die zweite Frage muß ebenfalls verneint werden, da es neben *inneren* auch *äußere* Gründe gibt. Äußere Gründe sind Gründe, für die es nicht *analytisch wahr* ist, daß sie zugleich Motiv des Handelnden sind. (Die erste Frage ist nicht nur für innere, sondern auch für äußere Gründe mit „nein" zu beantworten.)

Die dritte Frage gibt Anlaß, den Begriff des *„prima-facie"-Grundes* einzuführen. Es gibt Situationen, in denen Gründe für und gegen eine Entscheidung abgewogen werden müssen oder auch Gründe für h und Gründe für h'. Am Ende einer solchen Abwägung steht eine *wohlbegründete* Entscheidung für eine der offenstehenden Handlungsmöglichkeiten, z.B.: h. Dem Prädikat „wohlbegründet" entspricht ein Handlungsgrund G*, der nicht damit vereinbar ist, daß es für eine andere wohlbegründete Handlung h' einen entsprechenden Grund G* gibt. Das „*" soll kennzeichnen, daß es sich in diesem Fall nicht um bloße „prima-facie"-Gründe handelt.

Die übliche Interpretation des Bayes'schen Kriteriums besagt, daß eine rationale Person sich für h entscheidet, *weil* h den Erwartungswert der subjektiven Wünschbarkeit maximiert. T1 wäre in dieser Interpretation ein Modell der Begründung* einer Handlung. Der konsequentialistische Charakter dieser Interpretation von T1 wird deutlich, wenn wir den Begriff eines *„konsequentialistischen prima-facie-Handlungsgrundes"* einführen: G ist für P konsequentialistischer (innerer) prima-facie-Grund für h genau dann, wenn P einen Wunsch Z hat und P meint, daß Z durch h erreicht wird (Abschwächungen wären: „... gefördert wird" oder: „...wahrscheinlich – zumindest wahrscheinlicher – gemacht wird"). Diese Charakterisierung konsequentialistischer Handlungsgründe läßt sich dadurch trivialisieren, daß man jeder absichtlichen Handlung einen Wunsch zuordnet, nämlich den Wunsch, diese Handlung auszuführen. Ein Handlungswunsch im Sinne einer Absicht$_0$ ist jedoch begrifflicher Bestandteil

jeder Handlung; denn Handeln ist absichtliches$_o$ Tun. Diese Art von Wünschen muß daher in der konsequentialistischen Interpretation von T1 ausgenommen bleiben.

Aber auch Wünsche im Sinne *vorausgehender Absichten* kommen nicht infrage, denn die Erfüllungsbedingung vorausgehender Absichten ist – wie im letzten Abschnitt gezeigt wurde – die entsprechende Handlung selbst. Die Wünsche richten sich auf einen Zustand Z, Z ist Ziel der Handlung h: Z zu erreichen bildet eine *motivierende Absicht* der Handlung h. Der Handelnde nimmt an, Z sei eine (sichere, wahrscheinliche oder mögliche) *Konsequenz* der Handlung (der Wunsch kann sich auch auf Konsequenzen beziehen, die *Ergebnis* der Handlung h sind).

Dagegen kommt die Tatsache, daß h zu einem Handlungstyp T gehört, als ein mögliches Handlungsziel nicht infrage. Wenn P die Absicht hat, die Wahrheit zu sagen, dann hat P eine vorausgehende Absicht, die erfüllt wird, wenn die betreffende Sprachhandlung h zu dem Handlungstyp „eine wahre Aussage machen" gehört. Man kann dieser Absicht durchaus auch einen Wunsch zuordnen, der durch h erfüllt wird. Aber der Gegenstand dieses Wunsches ist von besonderer Art, er bezieht sich nicht auf den Zustand der Welt als Folge der Handlung, sondern er bezieht sich auf etwas, das mit der Handlung in einem logisch unauflöslichen Zusammenhang steht. Das Modell einer Handlung, die den weiteren Verlauf der Welt beeinflußt, ist hier nicht angemessen: Ob der Wunsch erfüllt wurde oder nicht, erkennt man nicht am Zustand der Welt, sondern daran, von welcher Art die Handlung ist. Zwischen der Art der Handlung und dem Zustand der Welt als Folge der Handlung gibt es jedoch keine eindeutige Zuordnung. Die auf Aristoteles zurückgehende Unterscheidung von intrinsisch und extrensisch Gutem weist in eine ähnliche Richtung. Wenn das, wonach der Handelnde strebt, nicht Folge der Handlung ist, sondern Charakteristikum der Handlung selbst, dann ist das erstrebe *Gute* Charakteristikum der Handlung und nicht eines Zustandes, der ggf. durch diese Handlung hervorgerufen wurde – die Handlung wäre damit intrinsisch gut.

Konsequentialistische Handlungsgründe entsprechen einem bestimmten Typ motivierender Absichten, wie sie im letzten Abschnitt eingeführt wurden. Dort hatten wir motivierende Absichten unterschieden, die auf eigene Handlungen gerichtet sind (genauer: deren Erfüllungsbedingungen Handlungen als vermutete Konsequenzen der begründeten Handlung sind) und solche, die nicht auf Handlungen gerichtet sind – wir haben diese „einfache" motivierende Absichten genannt. Konsequentialistischen Handlungsgründen entsprechen motivierende Absichten, die in keiner Weise auf Eigenschaften der betreffenden Handlung oder auf Eigenschaften anderer eigener Handlungen Bezug nehmen.

Für die konsequentialistische Interpretation von T1 repräsentiert die subjektive Bewertungsfunktion, deren Erwartungswert von einer rationa-

len Person maximiert wird, die unterschiedlichen konsequentialistischen prima-facie-Gründe für offenstehende Handlungen: In diese Funktion sind alle konsequentialistischen prima-facie-Gründe gegeneinander abgewogen und gewichtet eingegangen, sie faßt diese in einer kohärenten resultierenden Gesamtbewertung zusammen. *Die subjektive Bewertungsfunktion von T1 bildet den (konsequentialistischen) Grund* für* h: h hat für P den höchsten subjektiven Erwartungswert aller offenstehenden Handlungen.

T2 läßt sich dagegen nicht als ein Modell *innerer* – und a fortiori nicht als Modell *konsequentialistischer – Handlungsgründe* interpretieren. Keines der Rationalitätspostulate und keine Kombination dieser Postulate von T2 impliziert, daß eine rationale Person Handlungen unter dem Gesichtspunkt ihrer Folgen bewertet. Ob die Postulate von T2 erfüllt sind, hängt ausschließlich von der Gestalt der Präferenzrelation ab. Eine konsequentialistische Theorie der Handlungsrationalität läßt sich daher entgegen erstem Augenschein mit dem Theorem der Äquivalenz von T1 und T2 nicht rechtfertigen.

An dieser Stelle kann folgendes festgehalten werden:

1. Die Rationalitätstheorien T1 und T2 sind logisch äquivalent.

2. Die konsequentialistische Standardinterpretation von T1 wird durch T2 jedoch nicht gestützt. T1 und T2 sind äquivalent nur insofern, als die subjektive Bewertungsfunktion u von T1 als bloße *numerische Repräsentation* der im Sinne von T2 konsistenten individuellen Präferenzrelation gilt. Jede darüberhinausgehende inhaltliche Eingrenzung dieser Funktion u hebt die logische Äquivalenz von T1 und T2 auf.

3. Eine nicht-konsequentialistische Theorie der Entscheidungsrationalität muß entweder aufzeigen, daß mindestens eines der Postulate von T2 entbehrlich ist oder daß T1 mit einer nichtkonsequentialistischen Interpretation versehen werden kann.

Man muß sich darüber im klaren sein, daß jede Interpretation der Funktion u, für die u nicht eine bloße Metrisierung der zugrundeliegenden Präferenzen, sondern etwa Ausdruck der subjektiven Bewertung von Zuständen ist, der These, eine rationale Person maximiere den Nutzenerwartungswert, einen völlig anderen Status verleiht: Die Erwartungswert-Maximierungsthese läßt sich bei dieser Interpretation durch die Kohärenzpostulate von T2 nicht mehr rechtfertigen.

§ 10 Propositionale Einstellungen

Die Bayes'sche Entscheidungstheorie wird häufig so interpretiert, daß sie Handlungsrationalität auf zwei Arten propositionaler Einstellungen zurückführt: wünschen und glauben (annehmen, vermuten). So wird die subjektive Bewertungsfunktion über Zustände oder Propositionen bzw. Ereig-

nisse als Wünschbarkeit interpretiert, und die Wahrscheinlichkeitsfunktion repräsentiert die Neigung des Individuums, das Eintreffen eines (zukünftigen) Ereignisses anzunehmen.[3] Da Handlungen nach übereinstimmender Auffassung aller Ansätze philosophischer Handlungstheorie in der einen oder anderen Weise (sei es kausal, sei es logisch, sei es von einer besonderen „dritten" Art) mit Intentionen bzw. Absichten verknüpft sind, ergibt sich damit eine Theorie rationaler Absichten, die in dem Sinne reduktionistisch ist, als sie den Begriff der rationalen Absicht (als diejenige Absicht, die mit einer rationalen Entscheidung verknüpft ist) auf die Begriffe des Wünschens und des Glaubens reduziert.

Wenn in diesem Zusammenhang von Absichten die Rede ist, dann handelt es sich im Sinne der Unterscheidung von § 7 um *vorausgehende Absichten*. Insofern könnte man die Entscheidungstheorie als *Theorie rationaler vorausgehender Absichten angesichts gegebener subjektiver Wünsche und Vermutungen* ansehen. Eine vernünftige vorausgehende Absicht richtet sich auf die Handlung mit dem jeweils höchsten Erwartungswert, sollte diese nicht offen stehen, auf die mit dem nächsthöchsten Erwartungswert etc.[4] Die so entstehende Rangordnung spiegelt sich in den numerischen Werten der durch das Nutzentheorem gesicherten Funktion u.

Je nach Interpretation der Entscheidungstheorie als einer normativen, deskriptiven oder explikativen Theorie hat diese Reduktion den Charakter einer These mit normativem, deskriptivem bzw. begriffsexplikativem Charakter. Im ersten Fall lautet der Kernsatz der Theorie: Eine Person P *sollte* diejenigen vorausgehenden Absichten haben, die in der von der Theorie ausgesagten Weise mit ihren Wünschen und Annahmen verknüpft sind; im zweiten Fall: Eine Person *hat* diese vorausgehenden Absichten; und im dritten Fall wird der *Begriff* der vorausgehenden Absicht (in der von der Theorie vorgegebenen Weise) durch die Begriffe des Wünschens und des Annehmens expliziert bzw. idealiter definiert.

Die Entscheidungstheorie als reduktionistische Theorie der Rationalität aufzufassen setzt voraus, daß die zu optimierenden Bewertungsfunktionen (Nutzenfunktionen) tatsächlich Wünschbarkeitsfunktionen, also Bewertungsfunktionen sind, die die subjektiven Wünsche des Handelnden im

[3] Natürlich gibt es andere Möglichkeiten der Umschreibung und damit auch der Interpretation der entscheidungstheoretischen Grundbegriffe: So ließe sich etwa die Wahrscheinlichkeitsfunktion als die Annahme bestimmter objektiver Wahrscheinlichkeiten durch den Handelnden interpretieren – diese Interpretation macht aber immer schon dann Schwierigkeiten, wenn es sich um einmalige, nicht beliebig wiederholbare Ereignisse handelt. Komplexere Konzeptionen könnten die Wahrscheinlichkeitsfunktion, wie sie in das Bayes'sche Modell eingeht, als Repräsentanz subjektiver Vermutungen über vorliegende objektive Wahrscheinlichkeiten und/oder Annahmen der Person zu objektiven Wahrscheinlichkeiten interpretieren, wobei in letzterem Fall objektive Wahrscheinlichkeiten durchaus auch ohne eine frequentistische Interpretation aufrechterhalten werden können.

[4] Wenn zwei Handlungen den gleichen Erwartungswert und keine andere offenstehende Handlung einen höheren Erwartungswert hat, dann kann eine rationale Absicht auf jede der beiden Handlungen, aber nicht auf beide zugleich gerichtet sein – die rationale Person verhält sich nicht wie Buridans Esel.

Sinne innerer Gründe widerspiegeln. Für reale Personen ist diese Prämisse sicher häufig oder meist nicht erfüllt. Die wesentliche Frage ist jedoch nicht, ob diese Prämisse bei realen Personen je erfüllt ist, sondern ob anzunehmen ist, daß sich die Wünsche der (im Sinne der Theorie) idealen rationalen Person in den subjektiven Bewertungsfunktionen des Bayesianischen Modells widerspiegeln. Die bisweilen vertretene Ansicht, diese Frage sei nicht sinnvoll, da die Wünsche der Person ja erst durch das Konstruktionsverfahren der subjektiven Bewertungsfunktion über Testfragen etc. konstituiert werden, beruht auf einer zumindest für diesen Fall abwegigen operationalistischen Auffassung. Eine Theorie über den Zusammenhang von Wünschen und Handlungen muß sich vor dem Hintergrund unseres vortheoretischen Verständnisses als adäquat erweisen.

Die zweite kritische Frage an T1 in konsequentialistischer Interpretation ist: Wird der Zusammenhang zwischen Wünschen (und Glauben) der handelnden Person und rationalen Intentionen (vorausgehenden Absichten) adäquat widergegeben? Gibt es tatsächlich diesen engen Zusammenhang zwischen Wünschen, Annahmen und Intentionen, wie er vom Bayes'schen Konsequentialismus behauptet wird? Für die konsequentialistisch verstandene Theorie T1 ist die Entscheidung zwischen Handlungsoptionen eine Wahl des besten Mittels. Die Handlungsziele sind subjektive handlungsrelevante Wünsche und Neigungen, die sich in der konkreten Entscheidungssituation zu (handlungs-) motivierenden (konsequentialistischen, s. 3.2) Absichten verdichten: Die zu maximierende Funktion u repräsentiert daher die motivierenden Absichten der handelnden Person. Intrapersonell divergierende *handlungsrelevante* Wünsche werden gegeneinander abgewogen und in einer einheitlichen Bewertungsfunktion zusammengefaßt, die als quantitative Repräsentation motivierender Absichten aufgefaßt werden kann. Soweit dieser Weg nicht gangbar erscheint, da manche Wünsche *unvergleichbar* sind (etwa der Wunsch, das eigene Leben zu erhalten und eine aufregende Motorradfahrt zu unternehmen), erfolgt diese Abwägung über den Umweg fiktiven Entscheidungsverhaltens zwischen unsicheren Alternativen (Lotterien), also über T2. Die Faszination des Bayes'schen Konsequentialismus erklärt sich auch aus dieser Eleganz der Problemlösung (so „modell-platonistisch" sie sich ausnimmt): Die Kohärenz der Wünschbarkeiten wird über T2 hergestellt und schlägt sich in einer quantitativen Bewertungsfunktion nieder, die der rationalen Person zugeordnet werden kann.

Es wurde gezeigt, daß die konsequentialistische Interpretation von T1 aufgrund der Bedingungen von T2 nicht logisch zwingend ist, aber dies allein wäre kein hinreichender Grund, den Bayes'schen Konsequentialismus zu verwerfen. Aufzuzeigen bleibt, daß er als Rationalitätskonzeption inadäquat ist.

4. Kapitel: Abgrenzungen

Um die konsequentialistische Konzeption von Handlungsrationalität von anderen Konzeptionen abzugrenzen, kann man sich nicht auf die intuitive Eindeutigkeit von Formulierungen, wie „eine Handlung aufgrund ihrer Folgen gutheißen", „Handlungen als bloßes Mittel wählen" oder in ethischem Kontext: „Entscheidungen im Hinblick auf das außer-moralisch Gute treffen", verlassen. Solche Formulierungen bieten ein allzu weites Feld von Interpretationsmöglichkeiten. Einer ersten Eingrenzung dieses Feldes dienten die bisher dargestellten Elemente des Konsequentialismus[1]. Zum Abschluß dieses Kapitels werden vier Gesichtspunkte behandelt, die die Bestimmung des Konsequentialismus weiter präzisieren sollen.

§ 11 Der eigenorientierte Konsequentialismus

Das „klassische" Modell der Entscheidungstheorie von L. J. Savage[2] legt sich überraschenderweise in der Interpretation der Grundbegriffe eindeutig auf eine eigenorientiert-konsequentialistische Auffassung von Rationalität fest. Die für den Wert einer Handlung relevanten Umstände bestimmen im Savage-Modell die Konsequenzen der Handlung (Handlungen als Funktionen von Umständen in Konsequenzen), wobei jedoch Konsequenzen in einer ganz besonderen Weise, nämlich als *„subjektive Zustände der (handelnden) Person"*, interpretiert werden. Dieses entscheidungstheoretische Modell kann daher zur Charakterisierung einer bestimmten Version des Konsequentialismus herangezogen werden, die wir die *„eigenorientierte Konsequentialismus-Version"* (ek-Version, entsprechend „ek-besser" und „ek-rational") nennen. *Rational ist eine Entscheidung im Sinne des eigenorientierten Konsequentialismus genau dann, wenn keine andere Entscheidung offenstand, die den subjektiven Zustand der handelnden Person verbessert hätte.*

Es ist (ziemlich) unerheblich, welche Bewertungsmaßstäbe diese *besser*-Relation über die Menge der subjektiven Zustände prägen, in (fast) jedem Fall ist diese Version konsequentialistischer Rationalität alles andere als

[1] Diese Eingrenzung wird dadurch erschwert, daß Theoretiker, die dem konsequentialistischen Paradigma verpflichtet sind, dazu neigen, ihre Grundkonzeption zu trivialisieren („gemäß den eigenen Präferenzen handeln"), während sie in der konkreten Anwendung häufig eine sehr enge Interpretation konsequentialistischer Rationalität wählen. Ein Beispiel dafür ist die sogenannte „ökonomische Theorie der Politik", deren Rationalitätskonzeption sich „programmatisch" nicht von der der rationalen Entscheidungstheorie und der klassischen und neoklassischen Ökonomie unterscheidet, während sie bei der Analyse politischer Phänomene – etwa des Verhaltens von Parteien und Verbänden – ganz spezifische Wertfunktionen unterstellt. Erst diese empirisch verstandenen Wertprämissen machen den besonderen Charakter dieses Analyseansatzes aus, vgl. die „Klassiker" A. Downs, *Ökonomische Theorie der Demokratie*, Tübingen 1968 und M. Olson jr., *Die Logik kollektiven Handelns*, Tübingen 1968.

[2] *Foundations of Statistics*, New York 1954.

trivial. Wenn h eine für die Person P ek-rationale Handlung in Situation S ist, dann gibt es keine andere offenstehende Handlung h', so daß die subjektive Situation von P als Konsequenz von h' besser ist als die subjektive Situation als Konsequenz von h in S[3]. Wenn P eine Spende leisten will, um etwas gegen das Elend in der Dritten Welt zu tun, dann hat er ein bestimmtes Ziel (Elend zu mindern), und als rationale Person erwägt er, welches Mittel diesem Ziel am besten dient. Zur Wahl stehen eine Reihe von Organisationen, die – mit verschiedenen Maßnahmen – Projekte in der Dritten Welt unterstützen. Das Ziel der Entscheidung – die motivierende Absicht – ist offensichtlich nicht auf die Verbesserung der subjektiven Situation des Handelnden ausgerichtet. Das *Kriterium* der Wahl ist nicht eigenorientiert.

Dennoch schließt diese Handlungsorientierung nicht aus, daß die Person ihren subjektiven Zustand als Konsequenz der Entscheidung verändert – sie hat jetzt „ein gutes Gewissen", sie kann für eine gewisse Zeit ihren relativen Wohlstand unbeschwerter genießen. Offensichtlich ist dieser Aspekt der Handlungskonsequenzen jedoch nicht Grund (oder Bestandteil einer Begründung) der Handlung, er könnte bestenfalls als Bestandteil einer (kausalen) Erklärung berücksichtigt werden. Eine Handlung kann unter allen offenstehenden die besten subjektiven Konsequenzen haben, ohne daß die Verbesserung der subjektiven Situation Kriterium der Entscheidung ist. So könnte es sein, daß der reine Altruist de facto seine subjektive Situation optimiert. Dennoch wäre es inadäquat, dem Altruisten ek-rationales Verhalten zu unterstellen: die Verbesserung der subjektiven Situation ist nicht sein Handlungsmotiv[4].

§ 12 *Der Konsequentialismus als Begründungsmodell*

Um Unklarheiten zu vermeiden, sprechen wir von Konsequentialismus nur im Sinne eines *Begründungsmodells.* Daher genügt es nicht, die Rationalitätskonzeption des eigenorientierten Konsequentialismus durch die Darstellbarkeit als Savage-Modell zu charakterisieren. Nur solche Handlungen, die aufgrund *vorausgehender Absichten* vollzogen werden, kom-

[3] Es ist nicht ausgeschlossen, daß diese „besser-Relation" objektivistisch interpretiert wird, denn es sind objektive Maßstäbe der Vergleichung subjektiver Zustände denkbar, d.h. Maßstäbe, die vom Urteil des Handelnden unabhängig sind.

[4] Diese Unterscheidung scheint mir der zentrale Gedanke der Egoismus-Kritik J. Butlers zu sein, vgl. *Fifteen Sermons* (1726), hg. von T. A. Roberts (1970) sowie C. D. Broad, „Egoism as a Theory of Human Motives", *Hilbert Journal* 48 (1950), abgedruckt in: *Broad's Critical Essays in Moral Philosophy*, hg. von D. Cheney, London/New York 1971 sowie ders., *Five Types of Ethical Theory*, 1930, Kap. 3. In unserer Terminologie: Der psychologische Egoismus verwechselt eine kausale Folge der Erfüllung einer vorausgehenden Absicht mit dem Motiv des Handelns. Natürlich freue ich mich, wenn meine Hilfe zum Erfolg geführt hat, aber diese Freude war nicht Motiv meiner Entscheidung. Man kann das zu der These verdichten: Die Verbesserung meines subjektiven Zustandes ist fast immer nur Konsequenz und selten Motiv. Dort wo die Verbesserung meines subjektiven Zustandes zum unmittelbaren Handlungsmotiv, d.h. zur motivierenden Absicht, wird, stellt sich diese gewünschte Konsequenz des Handelns nur selten ein.

men als ek-rational in Betracht. Zentraler Gegenstand von Rationalitätskriterien sind Verhaltensweisen oder Handlungen *in ihrer Kombination mit motivierenden Absichten* (Absichten$_v$). Da die Erfüllung der vorausgehenden Absicht durch äußere Umstände vereitelt werden kann, ist es sogar angemessener, den *gesamten Komplex motivierender und vorausgehender Absichten* als eigentlichen Gegenstand der Beurteilung anzusehen[5], unbeschadet dessen, daß sich dieser Komplex auch für den Betrachter der ersten Art natürlich erst im Vollzug der Handlung bzw. in einem bestimmten beobachtbaren konkreten Verhalten manifestiert. Aufgrund der engen Verbindung von Absichten$_v$ und Handlungen bzw. Verhalten ist es naheliegend, auch von ek-rationalem *Handeln*, ja, sogar von ek-rationalem *Verhalten* zu sprechen. Dennoch sollte deutlich sein, daß dies nur als eine facon de parler zulässig ist; oder genauer: nur dann zulässig ist, *wenn die in Frage stehende konkrete Handlung Erfüllungsbedingung einer konsequentialistisch motivierten vorausgehenden Absicht ist.*

§ 13 *Intrinsisch wertvolle Handlungen*

Ein Typus der Kritik des Konsequentialismus richtet sich gegen die Vorstellung, die Entscheidung für eine Handlungsoption sei eine bloße *Mittelwahl*. In der Grundfassung des Konsequentialismus: Eine Handlung h ist nur dann konsequentialistisch rational, wenn der Handelnde ein Ziel hat, das er glaubt, durch h erreichen zu können. Differenziertere Fassungen berücksichtigen subjektive Wahrscheinlichkeiten und Zielkonflikte, was sich im Bayes'schen Konsequentialismus im Erwartungswert-Maximierungs-Kriterium und der Interpretation der subjektiven Bewertungsfunktion als Repräsentation motivierender Absichten niederschlägt.

Ist dieses instrumentelle Verständnis auch auf intrinsisch wertvolle Handlungen anwendbar? Diese Frage bräuchte uns nicht zu beschäftigen, wenn sich zeigen ließe, daß sich prima facie intrinsisch motivierte Handlungen als extrinsisch motiviert rekonstruieren ließen. Betrachten wir folgendes Beispiel: P möchte sich bei S bedanken (b) und schickt ihr daher einen Blumenstrauß. Sicherlich ist das Zuschicken des Blumenstraußes (z) extrinsisch motiviert: Es ist Mittel zum Zwecke des Bedankens. P bedankt sich dadurch, daß er die Handlung z vollzieht: b ist Folge (nicht Ergebnis!) von z. z mag eine rationale Mittelwahl sein, denn b wird unter fast allen Umständen Folge von zu sein – es sei denn, daß die Zustellung ausbliebe, so daß P in einem Land mit unsicheren Postwegen möglicherweise eine andere Handlung vollzogen hätte, um z zu erreichen. b war der *Grund* für z: P hatte die motivierende Absicht b und glaubte, daß z ein geeignetes Mittel sei, um b zu erreichen, deshalb tat P z. Soweit fügt sich dieses Bei-

[5] Nicht zufällig spricht man von rationaler *Entscheidungs*theorie und nicht von rationaler Handlungstheorie. Entscheidungen sind jedoch noch keine Handlungen, sie sind in der hier gewählten Terminologie mit vorausgehenden Absichten zu identifizieren.

spiel gut in das konsequentialistische Modell. Nun ist jedoch b offenbar ebenfalls eine Handlung von P und damit setzt sich die Kette der konsequentialistischen Rekonstruktion fort: Was war der Grund für b? P würde vielleicht antworten, der Grund sei, daß er auch in Zukunft auf S angewiesen ist, und er hoffe, daß sie ihm infolge von b auch in Zukunft wohlgesonnen bleibt (w). Damit wäre die konsequentialistische Analyse zu einem befriedigenden Abschluß gelangt; denn w ist keine Handlung von P. Natürlich kann man danach fragen, was denn w für P erstrebenswert mache, aber die konsequentialistische Analyse hat dort ihre Grenze, wo es nicht mehr um Entscheidungen und Handlungen geht. Es ist jedoch auch denkbar, das P auf diese Frage antwortet, er habe mit z kein darüber hinausgehendes Ziel verfolgt, sein einziges Motiv sei gewesen, sich bei S zu bedanken. In diesem Fall scheint es kein Ziel von P zu geben, für das b ein geeignetes Mittel sein könnte – b ist *intrinsisch motiviert*, oder anders formuliert: „b ist für P ein intrinsisches Ziel" oder „b ist für P intrinsisch wertvoll". Entgegen dem konsequentialistischen Dogma scheint b jedoch trotz Abwesenheit eines Handlungszieles (und der Überzeugung des Handelnden, daß b ein geeignetes Mittel ist, dieses Ziel zu erreichen) durchaus wohlbegründet, denn auf die erneute Frage nach dem Grund für b würde P vielleicht antworten: „S hat mir in der Vergangenheit häufig wertvolle Hilfe geleistet", und eine Antwort dieser Art scheint ein guter Grund zu sein, sich bei S zu bedanken.

Dieses Beispiel zeigt, daß es intrinsisch motivierte Handlungen gibt, die sich nicht in das Modell des konsequentialistischen Bayesianismus einfügen lassen. Es bleibt zu prüfen, ob das für alle *intrinsisch wertvollen*[6] Handlungen gilt – Handlungen, deren Wert sich nicht ausschließlich an ihrem „ergon"[7] bemißt. Handlungen haben nicht immer, aber häufig – unabhängig von ihren Folgen – einen bestimmten „Wert in sich", wobei „Wert" hier zunächst subjektivistisch verstanden werden kann. Der Vollzug der Handlung selbst kann z. B. Freude machen: Man ißt nicht, nur um seinen Hunger zu stillen, das Speisen selbst kann Freude bereiten. Im Rahmen einer konsequentialistischen Analyse geht der *intrinsische* Wert einer Handlung als Merkmal der Konsequenzen der Handlung in die zu maximierende Wertfunktion ein. Das sprachliche Pendant müßte in diesem Beispiel etwas unbeholfen lauten: „Man ißt, *um* Freude zu haben". Der Zu-

[6] Wer den Gebrauch von Begriffen wie „intrinsisch wertvoll" ablehnt, da er eine objektivistische Interpretation nahelegt, kann stattdessen auch den Ausdruck „intrinsisch motiviert" setzen wie z. B. R. Audi, „Acting for Reasons", *Philosophical Review* 95 (1986) 511–546 oder A. Mele, „Effective Reasons and Intrinsically Motivated Actions", *Philosophy and Phenomenological Research* 48 (1988) 723–731.

[7] Die Nikomachische Ethik des Aristoteles beginnt mit den Sätzen: „Jede techne („Kunst") und jede episteme („Wissen"/„Wissenschaft"), ebenso jede praxis („Handlung") und prohairesis („Entschluß") scheint irgendein Gut zu erstreben. Darum hat man mit Recht das Gute als dasjenige bezeichnet, wonach alles strebt. Es zeigt sich aber ein Unterschied in den tele („Zielen"): denn die einen sind energeiai („Tätigkeiten") und die anderen sind erga („Werke") außer ihnen" (1094a).

stand der „Freude" muß dann allerdings begrifflich von dem des Essens getrennt werden (also nicht die spezifische Freude, Schnitzel zu essen): das (konkrete) Essen als Mittel zum Zwecke der Freude. Während man jedoch sofort versteht, was es heißt, gerne Kalbsschnitzel zu essen, macht die Trennung der Handlung von ihrer adverbialen, subjektiv-wertenden Charakterisierung gewisse Schwierigkeiten, denn diese Art der Analyse führt zu schwer faßlichen abstrakten mentalen Zuständen: Ziel des Kalbsschnitzel-Essens ist ein bestimmter Zustand, der der Freude – per se. Ich bestelle nicht Kalbsschnitzel, weil ich gerne Kalbsschnitzel esse, sondern weil ich einen bestimmten Zustand zu erreichen suche, der Folge dieser Tätigkeit ist. Adverbiale Formulierungen zur Beschreibung bestimmter persönlicher Präferenzen sperren sich auffallend gegen eine Zerlegung in zwei konkrete Vorgänge – z. B. in den des Essens und den der Freude.

Eine konsequentialistische Analyse, die den *intrinsischen Wert* von Handlungen berücksichtigt, würde daher eine begriffliche Zerlegung jeder intrinsisch wertvollen Handlung einmal in einen wertneutralen Gegenstand der Entscheidung, und zum anderen in ein wertbehaftetes Ergebnis dieser Entscheidung verlangen. Dieses Verfahren ist äußerst künstlich und läuft im wesentlichen darauf hinaus, die intrinsische Motivation einer Handlung dadurch in die konsequentialistische Analyse einzubetten, daß eine neue Beschreibung der Handlung vorgenommen wird, wobei die Handlung „unter dieser Beschreibung" extrinsisch motiviert ist. Da dieses Verfahren das Problem der intrinsischen Motivation nur verschiebt, bleibt es bei der schon in § 9 skizzierten These: Intrinsisch wertvolle Handlungen können wohlbegründet sein. Es wäre daher unangemessen, sie aus dem Anwendungsbereich einer Theorie ratioalen Handelns auszuschließen. *Intrinsisch wertvolle Handlungen können jedoch nicht konsequentialistisch rational sein.*

§ 14 Regeln und Konsequenzen

Ein besonders interessanter Fall der Anwendung dieser Zerlegbarkeitsforderung ist die Erfüllung einer Regel. Wenn eine Person P in einer bestimmten Situation etwas äußert (A), was der Wahrheit entspricht, dann ist es eine *Konsequenz* dieser (Sprech-) Handlung h, daß P die Wahrheit gesagt hat, P also nicht gelogen hat: h erfüllt die Wahrhaftigkeitsregel WR. P könnte, nach dem Grund für h befragt, antworten: „Diese (konkrete) Handlung h erfüllt die Regel WR" oder kurz: „Ich wollte nicht lügen". Wenn ihm in der betreffenden Situation nur zwei Handlungen offenstanden, nämlich A oder $\neg$ A zu äußern (h'), dann hat er offenbar diejenige Handlung gewählt, die als einzige seinem Ziel – die Wahrheit zu sagen – entsprach. Die Frage ist: Hat P mit h konsequentialistisch rational gehandelt?

Betrachten wir zunächst den eigenorientierten Konsequentialismus. Es könnte sein, daß die Einhaltung der Regel WR die h *motivierende Absicht* für P war, P diese motivierende Absicht aber aufgrund der „übergeordneten" Absicht hatte, späteren Ärger zu vermeiden. Es könnte also sein, daß sich P für h entschieden hat, weil h bessere subjektive Konsequenzen für P erwarten ließ als h'. In diesem Fall hätte P ek-rational gehandelt. Es könnte jedoch auch sein, daß P keine weitere Antwort geben kann und darauf besteht, daß er mit der Einhaltung der Regel WR nichts Weitergehendes bezweckte. Hat P auch in diesem Falle – wir wollen annehmen, daß seine Aussage zutrifft – ek-rational gehandelt? Es könnte sein, daß P aufgrund von h Nachteile in Kauf nehmen muß, die er bei h' nicht hätte gewärtigen müssen, h wäre dann nicht ek-rational. Aber könnte h dann nicht nach wie vor *konsequentialistisch rational*, wenn auch nicht ek-rational sein?

Es ist wichtig zu sehen, *daß die konsequentialistische Rationalitätskonzeption völlig trivialisiert würde, wenn die bloße Regelkonformität einer Handlung h hinreichen würde, um h als konsequentialistisch rational zu charakterisieren, sofern nur diese Regelkonformität einer motivierenden Absicht des Handelnden entspricht.* Es ist erforderlich, den Konsequentialismus gegenüber dieser Trivialisierung abzugrenzen. Während die eigenorientierte Version dagegen, wie wir oben gesehen haben, zumindest dann immun ist, wenn sie als Begründungsmodell verstanden wird (Kriterium der Handlung h ist die subjektive Situation des Handelnden, die „Konsequenzen" im Savage-Modell), erfordert diese Abgrenzung für den Konsequentialismus in seiner unspezifischen Form einen Rückgriff auf das Modell aus I.1. Handlungen sind dort als Wahl einer Verzweigung an einem Verzweigungspunkt des Baumes charakterisiert worden. Die Äste des Baumes sind als Folgen von Weltzuständen aufzufassen.

Die Konsequenzen einer Handlung sind in diesem Modell nichts anderes als *eine spezifische Wahrscheinlichkeitsverteilung* über die Äste des durch die Handlung gewählten Zweiges. Nun sind wir oben nicht näher darauf eingegangen, was zu den Charakteristika eines Weltzustandes gehört, was eine ideale, d.h. vollständige Beschreibung eines Weltzustandes wäre. Lassen wir die konkrete Äußerung A aus dem obigen Beispiel auf einen Zeitpunkt t zusammenschrumpfen. Zu diesem Zeitpunkt erfolgt der Übergang von z_i zu z_{i+1}. P hat sich damit für den Zweig h und gegen den Zweig h' entschieden: Er hat einen Nachfolger-Zustand z_{i+1} festgelegt. Der Sachverhalt, daß A der Wahrheit entspricht, gehört nicht zu den Charakteristika von z_{i+1} und damit zu den Merkmalen des Astes, der die Folgen der Handlung symbolisiert. Dieser Sachverhalt charakterisiert die Handlung h, es ist eine Eigenschaft von h, daß A der Wahrheit entspricht. Da der Zustand als eine vollständige Konjunktion generischer Sachverhalte interpretiert wird, ist er nicht einmalig: Er könnte auch Element eines anderen Astes sein. Die Art und Weise des Übergangs von z_i zu z_{i+1}

kann daher nicht Bestandteil von z_{i+1} sein. Das ist auch intuitiv plausibel, denn soweit man für die Identität möglicher Welten überhaupt eine klare Intuition hat, sind zwei Weltzustände sicher nicht schon deshalb unterschiedlich, weil der eine aus seinem Vorgänger durch eine Lüge hervorgegangen ist und der andere nicht. Mit anderen Worten: Es sind zwei gleiche Weltzustände denkbar, die eine unterschiedliche Vorgeschichte haben, und ob die betreffende Person gelogen hat, gehört zu dieser Vorgeschichte.

Teil II:
Konsequentialismus in der Ethik

5. Kapitel: Strikter ethischer Konsequentialismus

§ 15 *Der ethische Konsequentialismus als Spezialisierung konsequentialistischer Rationalität*

Den *ethischen Konsequentialismus* kann man als eine Spezialisierung konsequentialistischer Rationalität verstehen. Die Spezialisierung bezieht sich zum einen auf den Inhalt der zu maximierenden Werte und zum anderen auf den Anwendungsbereich der Theorie. Statt den Konsequentialismus als eigene Theorie zu verstehen, kann man ihn mit der Menge derjenigen Theorien, deren Kriterium richtigen Handelns konsequentialistisch ist, identifizieren. Der ethische Konsequentialismus wäre dann die Menge der normativ-ethischen Theorien, deren Kriterium richtigen Handelns konsequentialistisch ist.

Diese Beziehung von ethischem und allgemeinem Konsequentialismus wird schon durch die Wahl der Terminologie nahegelegt, und deshalb muß betont werden, daß sie keineswegs trivial, sondern Ausdruck der folgenden *inhaltlichen* These ist: Wer moralisch richtig handelt, handelt zugleich rational. Aus dieser These ergibt sich das metatheoretische Postulat: Eine normativ-ethische Theorie richtigen Handelns muß sich in eine allgemeine Theorie rationalen Handelns einbetten lassen. Deshalb gilt: Wenn die konsequentialistische Rationalitätskonzeption adäquat wäre, dann müßte auch jede adäquate normativ-ethische Theorie konsequentialistisch sein. Die Umkehrung gilt natürlich nicht. Nun könnte es sein, daß der Konsequentialismus eine zutreffende Theorie rationalen Handelns in allen Entscheidungssituationen ist – ausgenommen diejenigen, welche eine moralische Abwägung verlangen. In diesem Falle wären bereichsspezifische Kriterien angemessen, womit eine generelle Theorie der Entscheidungsrationalität obsolet würde.[1] Daß der Weg bereichsspezifischer Handlungskriterien nicht gangbar ist, wird erst die genauere Analyse zeigen. Es wird sich dabei herausstellen, daß die zentralen Probleme der Entscheidungsrationalität einerseits und diejenigen der Moralität von Handlungen andererseits eine gleiche oder zumindest sehr ähnliche Struktur aufweisen, was eine kohärente normative Theorie rationaler und moralischer Entscheidung zumindest nahelegt.

Die einfachste Möglichkeit, konsequentialistische Rationalität und ethi-

[1] Diesen Weg hat z. B. P. Suppes eingeschlagen, vgl. „The Concept of Obligation in the Context of Decision Theory" in: *Science, Decision and Value*, hg. von J. Leach/R. Butts/G. Pearce, Dordrecht 1973.

schen Konsequentialismus in Beziehung zu setzen, ergibt sich durch ein normatives Kriterium, das sich ausschließlich auf die subjektiven Wertfunktionen als Repräsentation motivierender Absichten des Handelnden bezieht. Dieses Kriterium verlangt, daß die subjektive Wertfunktion des moralisch Handelnden unparteiisch ist. Unabhängig davon, wie dieses Kriterium konkret formuliert ist, legt es *eine einzige moralische Rangordnung möglicher Welten* fest. Diese Rangordnung ist nicht abhängig von der handelnden Person – sie ist invariant gegenüber den Umständen der Entscheidung, den Handlungsoptionen und den Handelnden. Eine Handlung ist moralisch richtig, wenn sie die quantitative Repräsentation dieser unparteiischen Rangordnung möglicher Welten (bzw. ihren Erwartungswert) maximiert. Wir nennen diese Fassung *„strikter ethischer Konsequentialismus"*.

Das Paradigma einer normativ-ethischen Theorie dieser Art ist der klassische (direkte) Handlungsutilitarismus. Da der Utilitarismus jedoch meist nicht nur durch die konsequentialistische Beziehung zwischen dem (unparteiisch) Guten und dem moralisch Gebotenen (dem Rechten) charakterisiert wird, sondern auch durch eine inhaltliche Bestimmung des Guten, sollte der strikte ethische Konsequentialismus, um Mißverständnisse auszuschließen, nicht mit dem Handlungsutilitarismus identifiziert werden.

§ 16 *Der utilitaristische Idealtypus*

Es gibt eine ganze Reihe von Spielarten des Utilitarismus, deren Details hier nicht wesentlich sind. Unter diesen Spielarten befinden sich Konzeptionen, die in dem hier dargelegten Sinne nicht konsequentialistisch sind. Neuerdings sind Theorien entwickelt worden, die man nicht einmal als teleologisch bezeichnen kann[2]. Das Prädikat „utilitaristisch" wird in der philosophischen Diskussion bisweilen im Sinne von „konsequentialistisch", oft aber auch im Hinblick auf eine spezifische Theorie des intrinsisch Wertvollen gebraucht[3]. Uns interessiert der Utilitarismus als Paradigma konsequentialistischer Ethik. Wir gehen daher von einer Version des Utilitarismus aus, die diesem Erfordernis genügt, und verschieben

[2] Vgl. R. Trapp, *Nicht-klassischer Utilitarismus*, Frankfurt/M. 1988; zur Terminologie s. u. § 25.

[3] So beruht für J. St. Mill der Utilitarismus im wesentlichen auf der Annahme, daß Lust und Freiheit von Unlust die einzigen Dinge sind, die als Endzwecke wünschenswert sind, vgl. *Utilitarianism*, London 1863, Kap. 2, während in der zeitgenössischen, überwiegend angelsächsischen Utilitarismus-Diskussion der konsequentialistische Charakter des Utilitarismus in den Mittelpunkt gerückt ist: „Act-Utilitarianism is the view that the rightness or wrongness of an action is to be judged by the consequences, good or bad, of the action itself", J. J. C. Smart, „An Outline of a System of Utilitarian Ethics", in: *Utilitarianism – For and Against*, Cambridge 1973, S. 8, ebenso D. H. Hodgons: „Utilitarian appraisal is appraisal by reference to consequences. The thesis that acts should be so appraised is the common feature of the moral systems, which come under the name ‚utilitarianism'", *Consequences of Utilitarianism*, Oxford 1967, S. 1.

Fragen der Typologie ethischer Theorien auf den letzten Abschnitt dieses Kapitels.

Stärker als andere Zweige der Philosophie legt die Moralphilosophie Wert darauf, daß sich ihre Begriffe und Theorien nicht allzuweit vom moralischen „common sense" entfernen. Diese methodische Orientierung zeigt sich bei den unterschiedlichen Denkern wie Aristoteles, Kant, Butler, Moore, Rawls und Mackie und wurde von Erstgenanntem auch methodologisch in der Konzeption der praktischen Wissenschaft als einer „Umriß-Wissenschaft" explizit gemacht. Es gibt allerdings prominente Ausnahmen zu dieser Regel. Zu ihnen gehören neben Platon insbesondere auch zahlreiche Utilitaristen. Letztere betonen i. d. R. den Unterschied zwischen einer Rekonstruktion verbreiteter moralischer Überzeugungen (den „doxai" im Sinne Platons) und der Entwicklung einer moralischen Theorie.[4]

Eine in vieler Hinsicht idealtypische und hier zur Illustration besonders geeignete Utilitarismus-Version hat R. F. Harrod entwickelt.[5] Sie verbindet einen strengen Empirismus mit einer subjektivistischen Komponente und hat den Anspruch, eine *Definition* der zentralen moralischen Begriffe zu geben.[6] In der utilitaristischen Tradition herrscht eine gewisse Konfusion, was den Anwendungsbereich des moralischen Prädikats angeht, da aus der These, daß man eine Handlung im wesentlichen im Hinblick auf ihre Konsequenzen beurteilen müsse, vielfach gefolgert worden ist, daß *Konsequenzen* für sich genommen moralisch gut bzw. schlecht seien. Harrod schlägt dagegen vor, ausschließlich *Handlungen* und nicht Konsequenzen als solche als „moralisch gut" zu bezeichnen.[7]

Das Prädikat „moralisch gut" taugt demnach nicht für Ziele, sondern nur für *Handlungen als Mittel.* Auch in nicht-moralischen Kontexten charakterisiert das Prädikat „gut" i. allg. Mittel (relativ zu bestimmten Zwekken), das gilt auch für Gegenstände des alltäglichen Gebrauchs. J. L. Makkie geht sogar soweit, eine gemeinsame Definition des Prädikates gut in außermoralischer und moralischer Verwendung vorzuschlagen: „x ist gut" bedeute nichts anderes als „x genügt den infrage stehenden Erfordernissen"[8]. Die Tatsache, daß die Philosophie seit ihren Anfängen meist anders verfahren ist und sich über das Gute an sich – z. B. über die Frage, ob Er-

[4] Besonders deutlich bei R. M. Hare, *Moral Thinking* ...

[5] Vgl. R. F. Harrod, „Utilitarianism revisited", *Mind* 45 (1936) 137–156. Diese Konzeption wendet sich besonders gegen Moore's Abschwächungen des klassischen Utilitarismus, sowohl was die Werttheorie als auch was seinen konsequentialistischen Charakter anbelangt.

[6] G. E. Moores Position, daß der zentrale moralische Begriff, das Prädikat „gut" (in moralischem Gebrauch), nicht definierbar sei, ist für den Utilitarismus nicht charakteristisch.

[7] „But from the view that an act should be regarded as including the whole state of affairs expected to result from a volition, and that the goodness of an act depends essentially on what that state of affairs is, it does not follow that the state of affairs may itself be regarded as good", R. F. Harrod, „Utilitarianism revisited", *Mind* 45 (1936) 137–156, S. 141.

[8] J. L. Mackie, *Ethik. Auf der Suche nach dem Richtigen und dem Falschen*, Stuttgart 1981, Kap. 2.

kenntnis etwas Gutes sei – Gedanken gemacht hat, ist für Harrod kein Gegenargument. Vielmehr spreche die Tatsache, daß trotz dieser langen Tradition diese Art der Verwendung des Prädikats „gut" keinen Eingang in die Alltagssprache gefunden hat, dafür, daß „gut" ausschließlich Mittel und keine Zwecke charakterisieren könne.

Harrod umgeht das Problem, konkret zu bestimmen, was als intrinsisch gut gelten könne – ein Problem, dem G. E. Moore mit der umstrittenen Entwicklung einer „idealen" utilitaristischen Werttheorie zu begegnen suchte[9]. Vielmehr beruft er sich auf die ursprüngliche Intention des Utilitarismus und *definiert* „gut" relativ zu empirischen Sachverhalten: Handlungen sind *moralisch relevant* genau dann, wenn sie die *Zielsetzungen* anderer Personen *affizieren*, und sie sind *moralisch gut* genau dann, wenn sie diese *Zielsetzungen fördern*. Es erfolgt keine Einschränkung dieser Charakterisierung des utilitaristischen Kriteriums auf eine bestimmte Klasse von Zielen; denn damit wäre die strenge empirische Prüfbarkeit einer Handlung als moralisch gut unvereinbar. Wenn es für eine gute Handlung nicht ausreichte, daß sie die Ziele anderer Personen (optimal) fördert, sondern wenn zusätzlich verlangt würde, daß sie nicht irgendwelche Ziele, sonder nur gute Ziele anderer Personen fördert, dann würde sich das Problem der moralischen Beurteilung auf die der Ziele verschieben. Es trifft zwar zu, daß man das Handlungs*ziel* einer Person dann schlecht nennen kann, wenn es mit den Zielsetzungen anderer Personen konfligiert. Dies sollte man aber lediglich als eine andere Formulierung dafür auffassen, daß eine *Handlung*, die als Mittel dient, das besagte Ziel zu erreichen, schlecht ist, weil sie die Ziele anderer Personen behindert.

Während Harrod das Prädikat „gut" in moralischer Verwendung auf die tatsächlichen individuellen Zielsetzungen bezieht, bezieht der klassische Utilitarismus „gut" auf die Befriedigung von Bedürfnissen. Zielsetzungen im Sinne motivierender Absichten haben nur wenige Lebenwesen, vielleicht der Mensch allein. Diese moderne Variante des Utilitarismus ist daher anthropozentrisch, während der klassische Utilitarismus in seiner Konsequenz alle *bedürftigen* und damit im wesentlichen alle *empfindlichen* Lebewesen umfaßt und insofern eine universalistische Tendenz hat.

Die Rolle, die der klassische Utilitarismus Bedürfnissen zuweist, erklärt seine Abhängigkeit von psychologischen Annahmen. Dieses enge Abhängigkeitsverhältnis hat das utilitaristische Denken seit seinen antiken Wurzeln begleitet. Tatsächlich haben sich die meisten Utilitaristen intensiv mit den angeborenen und erworbenen Bedürfnissen der Menschen auseinandergesetzt. Bei J. St. Mill und J. Bentham kann man sogar den Eindruck gewinnen, sie würden die utilitaristische Position als eine direkte logische Konsequenz einer psychologischen Theorie verstehen. Das läßt sich zwar systematisch nicht aufrechterhalten, aber es gibt in der Tat eine interes-

[9] Vgl. G. E. Moore, *Principia Ethica*, Cambridge 1903, Kap. 6.

sante logische Beziehung: Angenommen, jede erlebte Befriedigung würde auf lange Sicht das „Anspruchsniveau" der betreffenden Person ein wenig anheben, so daß das Integral des Befriedigungsniveaus über das ganze Leben letztlich immer in etwa gleich hoch bliebe, dann wäre die utilitaristische Bewertung von Handlungen gegenstandslos: Zwischen allen Handlungen bestünde moralische Indifferenz.

Die Utilitarismus-Variante Harrods ist selbstverständlich vereinbar damit, daß die alltäglichen moralischen Urteile im Sinne des Utilitarismus falsch sind. Diese utilitaristische Position beschränkt sich darauf aufzuzeigen, was die *Bedeutung der moralischen Ausdrücke* ist. Damit werden die moralischen Meinungsverschiedenheiten über das, was richtig und falsch ist, auf eine empirische Ebene verlagert: Unrichtige moralische Urteile beruhen nach utilitaristischer Auffassung weder auf unterschiedlichen moralischen Begriffen noch auf unterschiedlichen ethischen Theorien, sondern auf unterschiedlichen Meinungen darüber, was der Fall ist – bzw. speziell: welche Konsequenzen einzelne Handlungen haben.

§ 17 *Metaphysische Aspekte*

Es ist für den Utilitarismus typisch, daß Personen – aber auch ihre Verdienste, Fehler und Lebensziele – nur mittelbar, nämlich als kontingente Bedingungen der zu maximierenden Aggregationsfunktion (z. B. individuellen Glücks), eine Rolle spielen.

Die charakteristischen Schwierigkeiten des Utilitarismus mit Fragen der Gerechtigkeit oder der Integrität der Person haben ihren Ursprung darin, daß Personen in einem bestimmten Sinne als unwesentlich für die ethische Beurteilung angesehen werden. Den verschiedenen Spielarten des Handlungs- wie des Regelutilitarismus (einschließlich des ethischen Bayesianismus) ist die Orientierung der moralischen Beurteilung an Zufriedenheit, Glück oder Lust als solcher gemeinsam – und zwar unabhängig davon, wer Träger dieser Zufriedenheit (Glück, Lust) ist. Gäbe es eine ungewöhnliche Person, deren Zufriedenheit sich unbegrenzt steigern ließe, z. B. durch den Genuß einer bestimmten Droge, dann ist es unter bestimmten empirischen Bedingungen denkbar, daß die gesamte Produktivität einer Gesellschaft unter utilitaristischer Beurteilung ausschließlich dazu verwendet werden müßte, diesem Zufriedenheitsmonster zu dienen. Da es vermutlich ein solches Wesen nicht gibt, sind die kontraintuitiven Schlußfolgerungen des Utilitarismus aus kontingenten Gründen scheinbar irrelevant, trotzdem weisen sie auf ein Charakteristikum des Utilitarismus hin, nämlich daß eine bestimmte Größe (sei es Zufriedenheit, Glück, Lust o. a.) zum Kriterium des moralisch Richtigen wird, von der zwar anzunehmen ist, daß sie ohne menschliche (oder tierische) Träger nicht existieren kann, deren personale Verankerung jedoch unberücksichtigt bleibt.

Der Utilitarismus in seinen verschiedenen Varianten ist sicherlich mit einer Vielzahl metaphysischer Positionen vereinbar, aber es ist zugleich unverkennbar, daß er eine besondere Nähe zu einer *reduktionistischen Theorie der Person* aufweist, wie sie im indischen Kulturkreis von einigen buddhistischen Strömungen, in der Antike vermutlich von Heraklit, sicherlich von den Atomisten und Epikureern und in unserer Zeit z. B. von Grice und Parfit vertreten wird. Nach reduktionistischer Auffassung ist unter der sogenannten Identität der Person nichts anderes als eine lose Verbindung physikalischer und psychologischer Sachverhalte zu verstehen, die eine Person nicht als eine eigenständige reale Entität konstituieren – vielmehr beruht das, was wir als kontinuierliche Existenz einer Person über die Zeit empfinden, auf unserer Fähigkeit, frühere Ereignisse zu erinnern und aus Erfahrungen zu lernen.

Wie bei anderen nicht belebten Gegenständen der täglichen Erfahrung, etwa Steinen, Gläsern, Städten, sind die Identitätskriterien vage und letztlich rein konventionell: Ein Glas, aus dem ich eine Ecke entferne, ist immer noch dasselbe, unterdessen beschädigte Glas. Ob wir allerdings noch von demselben Glas sprechen, wenn es zu pulversiertem Glasstaub verarbeitet wurde, ist zweifelhaft. Ein Schiff, daß so grundlegend renoviert wird, daß nur noch ein Bruchteil des alten Materials in der neuen Version enthalten ist, kann man zwar als das gleiche, mittlerweile aus anderen Materialien bestehende Schiff bezeichnen, ebenso aber auch als ein ganz anderes Schiff, welches nach einem ähnlichen Bauplan entwickelt worden ist. Nach reduktionistischer Auffassung steht es nicht besser um die Identitätskriterien bei Personen. Die Frage, ob jemand etwa nach einer schweren Hirnverletzung die gleiche Person ist oder nicht, ist nach reduktionistischer Auffassung nicht eindeutig zu beantworten. In einem strengen ontologischen Sinne entspricht dem Begriff der Person nach reduktionistischer Auffassung keine reale Entität[10].

Für nicht-reduktionistische Konzeptionen läßt sich dagegen die Person nicht mit Systemen bestimmter psychologischer und/oder physiologischer Entitäten identifizieren[11]. Die Argumentation des Reduktionismus stützt sich auf verschiedene Paradoxien, in die sich unter bestimmten kontrafaktischen Annahmen jede nicht-reduktionistische Auffassung personaler Identität zu verwickeln scheint. Tatsächlich lassen sich leicht kontrafakti-

[10] In derselben Weise, wie es für einen methodologischen Individualisten keine Nationen oder Völker gibt, sondern nur einzelne Personen, die nach irgendeinem Kriterium begrifflich zu einer Nation zusammengefaßt werden. Über die einzelnen Personen hinaus und die Sachverhalte, die für bestimmte interpersonelle Beziehungen zutreffen, gibt es keine zusätzliche Entität, z. B. die Entität einer Nation. Vgl.: R. Tuomela, *A Theory of Social Action*, Dordrecht 1984, bes. Kap. 2 und 12.

[11] D. Parfit unterscheidet zwei nicht-reduktionistischer Ansätze: Der eine identifiziert die Person mit einer eigenständig existierenden Entität, etwa einer spirituellen Substanz, wie bei Descartes, der andere identifiziert die Person mit einem zusätzlich bestehenden Sachverhalt, was Parfit als die „further fact vie" bezeichnet; vgl. D. Parfit, *Reasons and Persons*, Oxford 1986, S. 210ff.

sche Beispiele konstruieren, für die unsere Intuitionen keine eindeutige Beurteilung erlauben. Wie wäre es z. B. zu beurteilen, wenn sich eine Person wie eine Amöbe verdoppeln könnte? Sicherlich wären die beiden entstandenen Personen zwei unterschiedliche Personen, aber wäre jede von ihnen nicht mit der Person vor dem Zeitpunkt der Teilung identisch? Wenn man diese Frage bejaht, dann ist die Transitivitätseigenschaft der Identitätsbeziehung verletzt. Das spricht in Parfits Augen dafür, personale Identität nur als eine facon de parler zu betrachten.[12]

Wenn Personen nichts anderes als lose Verknüpfungen von Empfindungen sind, es jedoch gute und weniger gute Empfindungen gibt, dann ist eine Werttheorie plausibel, die gute Empfindungen als solche zum Kriterium richtigen Handelns macht und die die Verteilung dieser guten Empfindungen auf Personen für irrelevant erklärt. Nach unseren moralischen Intuitionen macht es dagegen einen wesentlichen moralischen Unterschied aus, ob einem Kind ein bestimmter Verzicht auferlegt wird (z. B. in Form einer mühsamen Ausbildung), von dem es später als Erwachsener profitieren wird, oder ob der gleiche Verzicht dazu dient, einem anderen, z. B. einem mit dem Kind in keinerlei Beziehung stehenden Erwachsenen, diesen Vorteil zu verschaffen. Man stelle sich etwa vor, daß die zusätzliche Arbeitsleistung, die das Kind als späterer Erwachsener erbringt, einer anderen Person materiell zugute kommt. Die erste Art des Verzichts scheint – zumindest unter bestimmten Bedingungen – moralisch zulässig zu sein, während die zweite Art des Verzichts als unfair gegenüber dem Kind gelten muß.

Sicherlich kann es unter bestimmten Umständen ebenfalls unfair sein, dem Kind einen Verzicht aufzuerlegen, von dem es erst sehr viel später profitieren wird. Vielleicht würde man die Unfairneß einer solchen Verteilung von Zufriedenheit über die Lebensabschnitte mit dem Argument rechtfertigen, das eine gewisse Verwandtschaft mit einer reduktionistischen Auffassung aufweist: Wenn die Kindheit für sich betrachtet eine Phase mit eigenem Wert sei, dann dürfe sie nicht einfach zum Mittel eines erfolgreichen Erwachsenendaseins instrumentalisiert werden. Die Möglichkeit einer solchen Argumentation scheint sogar dafür zu sprechen, daß eine reduktionistische Metaphysik in der ethischen Theorie auch anti-utilitaristische Argumente stützen kann. Dies gilt besonders für das Phänomen der Zeitdiskontierung subjektiver Bewertungen, die i. d. R. von utilitaristischen Ethiken als unzulässig erachtet wird.[13] Diese anti-utilitaristischen

[12] Im Gegensatz allerdings zu den meisten Reduktionisten vertritt Parfit einen nicht-physikalischen Reduktionismus, d. h. diese Kontinuität der „Person" bestünde auch dann, wenn alle Zellen meines Körpers und besonders meines Gehirns und der Zustand zu einem bestimmten Zeitpunkt genau kopiert würden. Wenn das Original zerstört wird, dann handelt es sich nach Parfit bei der Kopie um dieselbe Person, vorausgesetzt, es besteht das, was er psychologische Kontinuität nennt. Es wäre in diesem Fall sinnvoll, meint Parfit, davon zu sprechen, daß ich diese Operation überlebt habe.

[13] Vgl. D. Birnbacher, *Verantwortung für zukünftige Generationen*, Stuttgart 1988.

Implikationen einer reduktionistischen Konzeption lassen sich jedoch zwanglos durch eine normative Annahme begründen, die nicht von einer reduktionistischen Theorie der Person abhängig ist: Unser Verständnis der Würde einer Person erlaubt die Instrumentalisierung ihrer Handlungen, ihrer Verhaltensweisen oder gar eines ganzen Lebensabschnittes auch intrapersonell nur unter besonderen Bedingungen. Die fraglos bestehenden Unterschiede der moralischen Qualität des intra- und des interpersonellen Falls lassen sich zwangloser damit erklären, daß das Kind im Erwachsenen fortlebt und die späte Gratifikation des Verzichtes zugleich eine Art Wiedergutmachung darstellt.

Für den strengen Reduktionisten gibt es über Empfindungen und Einstellungen hinaus nichts, was in der ethischen Beurteilung zu berücksichtigen wäre. Insofern ist es verständlich, daß mit dieser metaphysischen Position meist eine Vernachlässigung individueller Rechte und der Fragen der Gerechtigkeit einhergeht. Auch die Geschichte des ethischen Denkens zeigt, daß es eine enge Verknüpfung zwischen reduktionistischer Metaphysik und utilitaristischer Ethik gibt[14]. Eine reduktionistische Auffassung der Person als einer bloßen Folge von Ereignissen und speziell Empfindungen ist jedoch generell mit der Vorstellung einer moralisch handelnden Person schwer vereinbar. Denn jede Form rationaler Entscheidung, sei es das Verfolgen von Plänen, der Verzicht auf persönliche Vorteile oder eigenes Wohlergehen um spätere Vorteile willen oder zugunsten des Wohlergehens anderer, macht für so verstandene „Personen" letztlich wenig Sinn[15]. Eine reduktionistische Metaphysik stützt die spezifische Werttheorie des Utilitarismus, aber sie zerstört zugleich den minimalen Bestand anthropologischer Annahmen, die für eine jede ethische Theorie, oder genauer, für die praktische Relevanz einer jeden ethischen Theorie, unabdingbar sind.

[14] Vgl. dagegen die kritische Auseinandersetzung mit den behaupteten moralphilosophischen Implikationen einer reduktionistischen Sicht der Person von N. Daniels. „Moral Theory and the Plasticity of Persons", *Monist* 62 (1979) 265–287; vgl. auch: B. Williams, „Persons, Character and Morality", in: *The Identities of Persons*, hg. von A. Rorty, Berkeley, 1976.

[15] Vgl. dazu S. Scheffler, „Ethics, Personal Identity and Ideals of the Persons", *Can. J. Phil.* 12 (1982) 229–246.

6. Kapitel: Gerechtigkeitskonsequentialismus

§ 18 *Gerechtigkeit im strikten Konsequentialismus*

Ein häufig vorgebrachter und seit J. Rawls' *Theorie der Gerechtigkeit* als topos weithin akzeptierter Einwand gegen die utilitaristische Ethik verweist auf die unangemessene Berücksichtigung von Verteilungsfragen[1]. Eine konsequentialistische Theorie, die den Gerechtigkeitsaspekt als eigenständigen Parameter der moralischen Beurteilung berücksichtigt, nennen wir eine „g-konsequentialistische Theorie". Der G-Konsequentialismus versucht, den Aspekt der gerechten Verteilung über eine Veränderung der moralisch geforderten motivierenden Absichten des Handelnden zu berücksichtigen. Die vorausgehenden Absichten werden qua Bayes'schem Kriterium aus der diese motivierenden Absichten repräsentierenden subjektiven Wertfunktion abgeleitet.

Es ist für unsere Argumentation nicht wesentlich, welches spezifische Gerechtigkeitskriterium in diese Wertfunktion eingeht. Wir wählen daher das Rawls'sche Unterschiedsprinzip als Beispiel. Rawls nimmt an, daß Personen hinter einem Fairneß garantierenden „Schleier des Nichtwissens" im Urzustand zwei Prinzipien der Gerechtigkeit wählen würden, die die Grundstruktur der Gesellschaft prägen sollen: Das erste Prinzip verlangt gleiche maximale Freiheiten für jedes Mitglied der Gesellschaft, das zweite die Verteilung von Gütern in einer Weise, die für die am schlechtesten gestellte Gruppe der Gesellschaft am günstigsten ist[2]. Besonders das

[1] Vgl. M. McManus/G. M. Walton/R. B. Coffman, „Distributional Equality and Aggregate Utility: Further Comment", *Am. Ec. Rev.* 72 (1972) 489–496; A. K. Sen, „Rawls vs. Bentham: An Axiomatix Examination of the Pure Distribution Problem", in: *Reading Rawls*, hg. von N. Daniels, Oxford 1975; E. F. Becker, „Justice, Utility, and Interpersonal Comparisons", *Theory and Decision* 6 (1975) 472–484; P. F. Hammond, „Equity, Arrow's Conditions, and Rawls' Difference Principle", *Econometrica* 44 (1976) 793–804.

[2] J. Rawls gibt mehrere unterschiedliche Fassungen dieser beiden Gerechtigkeitsprinzipien an. Die Modifikationen sind jeweils Präzisierungen, während die ursprüngliche Formulierung der Prinzipien ihre intuitive Motivation deutlicher wiedergibt. „Ich behaupte, daß die Menschen im Urzustand zwei ganz andere Grundsätze wählen würden, einmal die Gleichheit der Grundrechte und -pflichten; zum anderen den Grundsatz, daß soziale und wirtschaftliche Ungleichheiten, etwa verschiedener Reichtum oder verschiedene Macht, nur dann gerecht sind, wenn sich aus ihnen Vorteile für jedermann ergeben, insbesondere für die schwächsten Mitglieder der Gesellschaft, vgl. *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975, § 3, S. 31f. Später in § 11: „1. Jedermann soll gleiches Recht auf das umfangreichste System gleicher Grundfreiheiten haben, das mit dem gleichen System für alle verträglich ist. 2. Soziale und wirtschaftliche Ungleichheiten sind so zu gestalten, daß (a) vernünftigerweise zu erwarten ist, daß sie zu jedermanns Vorteil dienen, und (b) sie mit Positionen und Ämtern verbunden sind, die jedermann offenstehen", § 11, S. 81. Das zweite Prinzip („Unterschiedsprinzip") wird nach einer ausführlichen Diskussion in § 13 noch einmal modifiziert zu: „Soziale und wirtschaftliche Ungleichheiten sind so zu regeln, daß sie sowohl (a) den am wenigsten Begünstigten die bestmöglichen Aussichten bringen als auch (b) mit Ämtern und Positionen verbunden sind, die allen gemäß der fairen Chancengleichheit offenstehen.", S. 104. Das Unterschiedsprinzip bezieht sich auf die Grundstruktur der Gesellschaft, und zwar vermittelt durch repräsentative Personen, deren Situation anhand eines Maßes der Grundgüter (vgl. § 15) bestimmt wird. Zur Modifikation des ersten Grund-

zweite Prinzip wird dem utilitaristischen Prinzip des größten Gesamtnutzens entgegengesetzt: Nachteile einer Person oder einer Gruppe von Personen dürfen nicht dadurch gerechtfertigt werden, daß sie zur Steigerung des Gesamtnutzens der Gesellschaft beitragen. Vorteile einzelner sind umgekehrt nur dann gerechtfertigt, wenn sie allen und besonders den schlechter gestellten Personen nützen. Um eine eindeutige Rangordnung der gesellschaftlichen Zustände zu erhalten, geben wir dem Unterschiedsprinzip folgende Fassung: Wir numerieren die repräsentativen Personen der Gesellschaft nach ihrer Ausstattung mit Grundgütern: p_1 sei die am besten gestellte Person und p_n die am schlechtesten gestellte Person eines bestimmten gesellschaftlichen Zustandes. Dann ist ein gesellschaftlicher Zustand z v-gerechter[3] als ein anderer gesellschaftlicher Zustand z', wenn p_n in z besser gestellt ist als in z' oder, falls diese Bedingung nicht erfüllt ist, wenn p_{n-1} in z besser gestellt ist als in z' oder, falls sie in beiden Zuständen gleich gestellt ist, wenn p_{n-2} in z besser gestellt ist als in z' etc. Diese Fassung hat den Vorteil, daß eine Rangordnung der gesellschaftlichen Zustände auch dann besteht, wenn die kausale Verkoppelung der sozialen Situation der unterschiedlichen Gruppen der Gesellschaft eher lose ist[4]. Dennoch berücksichtigt das Unterschiedsprinzip auch in dieser Fassung vorrangig die Interessen der am schlechtesten gestellten repräsentativen Person.

Betrachten wir zunächst eine Teilmenge aller gesellschaftlichen Zustände, die die Eigenschaft hat, daß jeweils zwei gesellschaftliche Zustände dieser Menge ein unterschiedliches Nutzenniveau der am schlechtesten gestellten repräsentativen Person aufweisen. Die Rangordnung innerhalb dieser Teilmenge aufgrund des Unterschiedsprinzips ist eindeutig: Da ein gesellschaftlicher Zustand z v-gerechter gegenüber einem gesellschaftlichen Zustand z' ist, wenn die am schlechtesten gestellte repräsentative Person in z besser gestellt ist als in z', ist die Rangordnung unter dem Aspekt der Verteilungsgerechtigkeit mit der Rangordnung der Zustände nach dem Nutzenniveau der jeweils am schlechtesten gestellten repräsentativen Person identisch. Als kardinales Maß der V-Gerechtigkeit bietet sich das absolute Nutzenniveau der jeweils am schlechtesten gestellten repräsentativen Person an. Die ideale moralische Person im Sinne des Unterschiedsprinzips maximiert den Erwartungswert dieser Funktion. Wenn man sich nicht auf Teilmengen dieser Art beschränken will, ergibt sich

Fortsetzung Fußnote von Seite 61

satzes vgl. § 39. Eine endgültige Fassung unter Berücksichtigung des gerechten Spargrundsatzes und einer präziseren Fassung der Vorrangregeln – der Freiheit vor der Verteilung sozialer und wirtschaftlicher Güter und der Verteilungsgerechtigkeit vor Effizienz – erfolgt in § 46, S. 336f.

[3] „v" steht hier für „*Verteilungs*gerechtigkeit", uneingeschränkt gerecht ist ein Zustand erst dann, wenn das erste Prinzip der gleichen maximalen Freiheit ebenfalls erfüllt ist und beide Prinzipien in einer „demokratischen" Interpretation verwirklicht sind, vgl. Rawls (1975) §§ 12–14, sowie Kap. 4.

[4] Vgl. A. K. Sen, *Collective Choice and Social Welfare*, London u. a. 1970, S. 135–141.

zwar das mathematische Erfordernis, transfinite Werte zuzulassen, um die lexikographische Vorordnung der jeweils schlechter gestellten repräsentativen Person aufrechtzuerhalten, aber im großen und ganzen scheint die konsequentialistische Einbettung von Gerechtigkeitsgesichtspunkten in diesem Beispiel gelungen zu sein.

Das Beispiel hat zwei Vorzüge, es ist einfach, und es knüpft an die wohl prominenteste Gerechtigkeitstheorie der Gegenwartsphilosophie an, die als Alternative zur utilitaristischen Sozialethik entwickelt wurde. Man kann sich aufgrund dieses Beispieles leicht vor Augen führen, wie eine Inkorporierung beliebiger Gerechtigkeitskriterien in eine konsequentialistische Theorie zu bewerkstelligen wäre. Reine Theorien der Verteilungsgerechtigkeit sind dabei einfacher zu handhaben als Theorien, die eine Abwägung von Verdienst-, Effizienz- und Gerechtigkeitskriterien versuchen. Daß aber selbst diese komplizierten Fälle sich einer konsequentialistischen Rekonstruktion nach der Art des obigen Beispiels nicht verweigern, zeigt eine erst jüngst entwickelte Theorie von R. Trapp[5].

§ 19 *Deontologischer Konsequentialismus*

Die Möglichkeit der konsequentialistischen Rekonstruktion läßt den Übergang von traditionellen utilitaristischen Theorien zu g-konsequentialistischen Theorien als fließend erscheinen. Dieser Eindruck täuscht jedoch: g-konsequentialistische Theorien sind im Gegensatz zu traditionellen utilitaristischen Theorien *nicht teleologisch*. Unter einer *teleologischen* Theorie verstehen wir eine normativ-ethische Theorie, für die das Gute außermoralisch und unabhängig vom Rechten ist[6] und das Rechte das

[5] R. Trapp, *„Nicht-klassischer" Utilitarismus: Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1988. Der von R. Trapp entwickelte „Gerechtigkeitsutilitarismus" berücksichtigt tatsächlich neben dem Prinzip des Subjektivismus (der moralische Wert einer Handlung ist Funktion der Nutzenniveaus der Betroffenen), dem Prinzip der Anonymität und der Nutzensteigerung (ein Zustand z ist dann höher zu bewerten als z', wenn das Maß der Ungleichheit in z und z' gleich, aber die Nutzensumme in z größer ist) sowie einigen weiteren Bedingungen, wie Pareto-Superiorität, Symmetrie und Ordnungseigenschaften, auch ein Prinzip der gerechten Nutzenverteilung (ein Zustand z ist höher zu bewerten als z', wenn z und z' die gleiche Nutzensumme haben, aber z eine geringere Ungleichheit der individuellen Nutzenniveaus aufweist) und ein Prinzip des Verdienstes, das sich in unterschiedlichen Graden der Berücksichtigung der individuellen Nutzenzuwächse in der Aggregation niederschlägt, S. 300–408.

[6] Das einflußreiche Buch von W. K. Frankena *Ethics*, Englewood Cliffs 1963, dt. *Analytische Ethik*, München 1972 führt den Begriff der teleologischen Theorie unter Rückgriff auf den Begriff des „außermoralischen" Wertes ein, während J. Rawls den Begriff der teleologischen Theorie ausschließlich durch die Unabhängigkeit des Guten vom Rechten charakterisiert. Frankena: „Eine teleologische Theorie behauptet, daß das grundlegende Kriterium dafür, was moralisch richtig, falsch, verpflichtend usw. ist, der außermoralische Wert ist, der geschaffen wird. Danach muß man sich, ob direkt oder indirekt, letzten Endes auf die vergleichsweise Summe guter Konsequenzen berufen oder vielmehr auf das vergleichsweise Übergewicht von guten gegenüber schlechten Konsequenzen", S. 32. Rawls: „Die einfachste Beziehung scheinen teleologische Theorien herzustellen: Das Gute wird unabhängig vom Rechten definiert, und dann wird das Rechte als das definiert, was das Gute maxi-

Gute maximiert. Eine *deontologische* Theorie ist eine normativ-ethische Theorie, für die das Gute nicht unabhängig vom Rechten ist. Damit wird nicht behauptet, daß jede nicht-teleologische Theorie deontologisch ist, da es ja normativ-ethische Theorien geben könnte, für die weder das eine noch das andere gilt. Dies scheint für einige Tugend- und Motivationsethiken[7] der Fall zu sein.

Für eine g-konsequentialistische Theorie hat Gerechtigkeit keinen ausschließlich instrumentellen Status: Ob ein gesellschaftlicher Zustand gerecht ist oder nicht, spielt nicht nur insofern eine Rolle, als dieser Sachverhalt das Maß des insgesamt verwirklichten (außermoralischen) Guten kausal beeinflußt (wie es für utilitaristische Theorien der Fall ist), sondern prägt unmittelbar den moralischen Wert eines gesellschaftlichen Zustandes. Man könnte auch sagen, das moralische Kriterium der Gerechtigkeit hat in g-konsequentialistischen Theorien einen *intrinsischen Wert*. Damit geht jedoch ein wesentliches Merkmal teleologischer Theorien verloren: Das Gute ist nicht mehr *vor aller Moral* „gegeben", sondern das Gute ist selbst Gegenstand moralischer Beurteilung.

Für Rawls ist seine Theorie der Gerechtigkeit deshalb deontologisch, weil die Verteilung der Güter selbst wiederum „als ein Gut genommen wird" und das Verteilungsproblem „nach intuitivem Verständnis (...) unter den Begriff des Rechten fällt"[8]. Tatsächlich ist dieses Argument nur auf die Vorordnung des Freiheitsgrundsatzes bei Rawls anwendbar; denn dieses Prinzip ist ein *Verpflichtungsgrundsatz* und gehört somit eindeutig zur Theorie des Rechten. Der moralische Wert einer Verteilung kann dagegen unabhängig von Verpflichtungsfragen beurteilt werden – zumindest ist das die Intention der meisten Gerechtigkeitstheorien einschließlich der von J. Rawls. Erst auf der Grundlage einer Theorie der gerechten Verteilung ergibt sich die Theorie des richtigen Handelns, die bei J. Rawls als Konsequenz der Konzentration auf Institutionen allerdings nur eine untergeordnete Rolle spielt[9]. Um Konfusionen dieser Art zu vermeiden, ist

Fortsetzung Fußnote von Seite 63

miert", S. 42. Tatsächlich ist es aus folgendem Grund sinnvoll, beide Charakterisierungen zu kombinieren: Da die Abgrenzung von „außermoralisch" nicht sehr scharf ist, ist es nach den üblichen Sprachgebrauch nicht ausgeschlossen, daß es etwas „außermoralisch" Gutes gibt, das nicht unabhängig vom Rechten ist. Ein Fall könnte in obigem Beispiel gesehen werden: Sicher ist das Wohlergehen der am schlechtesten gestellten Person etwas außermoralisch Gutes, dennoch versteht sich Rawls' Theorie als eine Maximierungstheorie dieser Größe nicht als teleologisch; denn diese Forderung ist Konsequenz eines Gerechtigkeitskriteriums und damit nicht unabhängig von der Theorie des Rechten. Andererseits könnte man die Beurteilung einer Güterverteilung unter dem Aspekt der Verteilungsgerechtigkeit unabhängig von einer Theorie des Rechten vornehmen, womit die entsprechende Theorie der Maximierung der Verteilungsgerechtigkeit ebenfalls noch nicht teleologisch wäre.

[7] Vgl. z. B. A. Mac Intyre, *Der Verlust der Tugend. Zur moralischen Krise der Gegenwart*, Frankfurt/M. 1987.

[8] Rawls (1975), S. 43.

[9] R. Nozicks Kritik an Gerechtigkeitstheorien vom Typ der „timeslice theories" beruht auf dem Argument, das Rechte (hier das, wozu Menschen berechtigt sind, welche unverletzba-

es zweckmäßig, zwei Einteilungen normativ-ethischer Theorien – die Einteilung in teleologische und nicht-teleologische Theorien einerseits und die Einteilung in konsequentialistische und nicht-konsequentialistische andererseits – auseinanderzuhalten.[10]

Normativ-ethische Theorien haben i. allg. zwei Bestandteile: einen Wertbestandteil und einen Verpflichtungsbestandteil. In teleologischen Theorien ist der Wertbestandteil primär und der Verpflichtungsbestandteil instrumentell, während in deontologischen Theorien der Verpflichtungsbestandteil primär ist und der Wertbestandteil (soweit überhaupt vorhanden) in Abhängigkeit von dem Verpflichtungsbestandteil bestimmt wird. Eine andere Formulierung dieser Dichotomie wäre: Deontologische Theorien sind solche, die Handlungen (Handlungsregeln oder generischen Handlungen) einen intrinsischen Wert zuweisen, während für teleologische Theorien eine Handlung niemals intrinsisch (moralisch) wertvoll (oder wertlos) ist, sondern ihren Wert nur durch ihren Beitrag zum verpflichtungsunabhängigen Guten erhält.

Fortsetzung Fußnote von Seite 64

ren Rechte sie haben) sei jeglichen Erwägungen der Verteilungsgerechtigkeit vorgeordnet: Eingriffe in individuelle Rechte unter dem Aspekt der Verteilungsgerechtigkeit seien unzulässig. Diese Kritik kann man daher auch als eine Radikalisierung der Rawls'schen Utilitarismuskritik ansehen: Während Rawls das Rechte nicht als bloßes Instrument des Guten begreift, aber Fragen der Verteilungsgerechtigkeit als Bestandteil der Theorie des Rechten betrachtet, kritisiert Nozick den verbliebenen „teleologischen" Bestandteil der Rawls'schen Theorie, wie er sich in der vom Rechten unabhängigen Beurteilung der Gerechtigkeit eines gesellschaftlichen Zustandes niederschlägt, vgl. R. Nozick, *Anarchy, State and Utopia*, New York 1974, dt. *Anarchie, Staat, Utopia*, München 1976, Kap. 7.

[10] Näheres dazu in Kap. 8.

7. Kapitel: Abgeschwächter Konsequentialismus

Für den strikten Konsequentialismus gibt es nur eine einzige moralisch zulässige Rangordnung möglicher Welten. Diese Rangordnung ist nicht abhängig von der handelnden Person, sondern invariant gegenüber den Umständen der Entscheidung, den Handlungsoptionen und den Handelnden. Eine Handlung ist moralisch richtig, wenn sie die quantitative Repräsentation dieser unparteiischen Rangordnung möglicher Welten (bzw. ihren Erwartungswert) maximiert. Auch der Gerechtigkeitskonsequentialismus stellt eine Form des strikten Konsequentialismus dar.

Begründet wird die Invarianz der Bewertung des strikten Konsequentialismus damit, daß jede normativ-ethische Theorie unparteiisch sein sollte und Wertungen nicht auf die Interessen oder persönlichen Umstände des Handelnden relativieren dürfe. Diese Forderung wird erstaunlicherweise sogar von streng subjektivistischen Ethiken übernommen[1]. Erstaunlich ist diese Forderung deshalb, weil die subjektivistische Werttheorie die zu maximierenden Werte in Abhängigkeit von gegebenen subjektiven Präferenzen bestimmt: D.h., es wird nicht der Anspruch objektiv gegebener (wenn auch nicht-moralischer) Werte erhoben, wie es eine „unparteiische", interpersonell invariante Wertfunktion nahelegen würde. Dennoch wird von dem moralisch Handelnden verlangt, daß er alle subjektiven Präferenzen *völlig gleichrangig* behandelt. Im strikten Konsequentialismus, auch dem subjektivistischer Art, kommt ein erstaunlicher Rigorismus zum Ausdruck, der von der moralisch handelnden Person „übermenschliche" Distanz zu den eigenen Wünschen und Interessen sowie zu den Wünschen und Interessen nahestehender Personen verlangt. Dieser Rigorismus ist in doppelter Hinsicht strenger als der Rigorismus einer deontologischen Theorie vom Kantischen Typ: Alle Handlungen werden gleichermaßen der moralischen Beurteilung unterzogen; denn für den strikten Konsequentialismus steht jede Handlung unter dem Aspekt der Folgenoptimierung. Darüber hinaus läßt der strikte Konsequentialismus (hinreichende Meßgenauigkeit vorausgesetzt) keinen Spielraum, da es i. allg. jeweils nur *eine* optimierende Handlung gibt, während selbst eine deontologische Theorie vom Kantischen Typus sich auf ein *Ausschließungs*kriterium beschränkt, das jeweils ganze Klassen von generischen Handlungen als *zulässig* erachtet und zudem nicht alle Handlungen einer moralischen Bewertung unterzieht.

Der strikte Konsequentialismus ist nicht die einzige Möglichkeit, der Forderung nach Unparteilichkeit gerecht zu werden, auch strikt-deonto-

[1] Im Falle des ethischen Bayesianismus schlägt sich das allerdings im Auseinanderklaffen „subjektiver" und „ethischer" Präferenzen nieder, was genau besehen das Eingeständnis enthält, daß es die moralisch handelnde Person nicht geben kann, vgl. Verf., *Entscheidungstheorie und Ethik*. München 1987, S. 105ff.

logische Theorien können so beschaffen sein, daß sie der besonderen Berücksichtigung persönlicher Neigungen und eigenorientierter Wertungen keinen Raum lassen. Vielfach wird jedoch jede Abschwächung der Unparteilichkeit als eine Abwendung vom konsequentialistischen Paradigma verstanden. Wir können uns dieser Auffassung nicht anschließen; denn sie verdeckt den engen logischen Zusammenhang zwischen einer konsequentialistischen Rationalitätskonzeption und ihrer Spezialisierung in der ethischen Theorie.

Im folgenden wird das Postulat der Unparteilichkeit der zu maximierenden Wertfunktion aufgegeben, ohne damit zugleich das konsequentialistische Entscheidungsprinzip in Frage zu stellen: Wir betrachten den Typus einer konsequentialistischen ethischen Theorie mit *interpersonell varianten* Bewertungen.

§ 20 *Erweiterte Sympathie*

Für eine konsequentialistische Theorie steht die Bestimmung einer Wertfunktion im Mittelpunkt, da sich die Verpflichtungsurteile aus der Maximierungsforderung ableiten lassen. Eine ausgereifte konsequentialistische Theorie enthält eine *quantitative* Wertfunktion. Da der Übergang von einer (qualitativen) besser-Relation zu einer quantitativen Wertfunktion jedoch unter der Voraussetzung bestimmter Bedingungen möglich ist[2], ist es zulässig, sich zunächst auf das Problem der *Rangordnung unter ethischen Aspekten* zu beschränken. Wir machen uns dafür zunächst das Konzept der „erweiterten Sympathie" zunutze, das in P. Suppes' vieldiskutiertem „grading principle of justice"[3] zum ersten Mal Verwendung gefunden hat. Das Konzept der „erweiterten Sympathie" wird hier auch deshalb eingeführt, weil es darüber hinaus wichtige Aufschlüsse über die Struktur subjektivistisch-konsequentialistischer Theorien gibt. Zunächst sind einige formale Vorbereitungen erforderlich.

1) Eine Gruppe von Personen K, deren Präferenzen entweder als unmittelbar Entscheidungsbeteiligte oder als – u. U. nur indirekt – Betroffene für die Bildung der ethischen Präferenzen relevant sind.

2) Eine Menge von Alternativen X, die in eine ethische Rangordnung gebracht werden sollen. Diese Alternativen, die man sich als eine hinreichend detaillierte Beschreibung eines *gesellschaftlichen Zustandes* vorstellen kann, enthalten jeweils eine Reihe von Positionen, die von einzelnen Individuen aus K eingenommen werden können. Da diese sehr allgemeine Interpretation der Entscheidungsalternativen in K. Arrows bahnbrechen-

[2] Vgl. § 8.
[3] Vgl. P. Suppes, „Some Formal Models of Grading Principles", *Synthese* 6 (1966), 284–306.

dem Werk[4] zur Theorie kollektiver Entscheidungen[5] verwendet wird, nennen wir die Elemente von X auch *Arrow-Alternativen*.

3) Die *Menge der Positionen*. Der Ausdruck „Positionen" könnte mißverständlich sein: Man darf sich die einzelnen Positionen nicht als Bestandteile einer hierarchischen Ordnung denken – das ist nur eine und eine eher uninteressante Möglichkeit –, vielmehr ist eine *Position* p eine anonyme Beschreibung aller für eine Person in einer bestimmten gesellschaftlichen Situation relevanten Umstände.

4) Eine *Positionszuordnung* φ gibt an, welches Individuum welche Position in x einnimmt: φ: $K \ni i \rightarrow p \in z$, $z \in Z$.

5) Wenn alle gesellschaftlichen Positionen einer Alternative bekannt sind, aber nicht, welches Individuum in welcher dieser Positionen ist, so nennen wir die Beschreibung dieser gesellschaftlichen Positionen einer Alternative einen *anonymen gesellschaftlichen Zustand* z.

DEF.: Sei P die Menge aller gesellschaftlichen Positionen, dann ist Pot(P) die *Menge aller anonymen gesellschaftlichen Zustände* Z. Die Angabe einer vollständigen Positionszuordnung führt – sozusagen qua Deanonymisierung – z in einen „gesellschaftlichen Zustand" x im Sinne Arrows über.

DEF.: Das geordnete Paar $\langle \varphi, z \rangle = x$ nennen wir daher einen *Arrowschen Zustand*. Die Menge aller Arrowschen Zustände sei X. Natürlich muß gefordert werden, daß die Anzahl der Positionen $\#z$ eines anonymen gesellschaftlichen Zustandes z mindestens so groß ist wie die Anzahl n der Personen in K. Wenn Z die Menge der anonymen gesellschaftlichen Zustände ist und Φ die Menge aller möglichen Positionszuordnungen, dann kann das Kreuzprodukt $\Phi \times Z$ mit der Alternativenmenge X identifiziert werden. Einem einzigen anonymen gesellschaftlichen Zustand z entsprechen $\#z!/(\#z-n)!$ gesellschaftliche Zustände, die durch Deanonymisierung aus z hervorgehen. Wenn man vereinfachend annimmt, daß die Anzahl der Positionen jedes anonymen gesellschaftlichen Zustandes gleich der Anzahl n der Personen aus K ist, dann entsprechen einem einzigen anonymen gesellschaftlichen Zustand z $n!/(n-n)! = n!$ gesellschaftliche Zustände – die Anzahl der gesellschaftlichen Zustände ist gleich der An-

[4] Vgl. K. J. Arrow, *Social Choice and Individual Values*, New York 1951, [2]1963 sowie „Values and Collective Decision Making", in: *Philosophy, Politics, and Society, Third Series*, hg. von P. Laslett u. W. G. Runciman, Oxford/New York 1967 und „Public and Private Values", in: *Human Values and Economic Policy*, hg. von S. Hook, New York 1967.

[5] Die nach wie vor beste zusammenfassende Darstellung der Theorie kollektiver Entscheidungen bietet A. K. Sen, *Collective Choice and Social Welfare*, San Francisco u.a. 1970. Neueren Datums, aber sehr knapp ausgefallen ist das Kapitel 22 „Social Choice Theory" von A. K. Sen in: *Handbook of Mathematical Economics*, Bd. III, hg. von K. Arrow u. M. Intriligator, Amsterdam u.a. 1986, S. 1073–1182 – es enthält zahlreiche Hinweise auf weiterführende Literatur. Eine kritische Betrachtung der methodologischen und philosophischen Aspekte der Logik kollektiver Entscheidungen aus unterschiedlichen Blickwinkeln gibt *Foundations of Social Choice Theory*, hg. von J. Elster u. A. Hylland, Cambridge u.a. 1986, darin bes. die Beiträge von A. Gibbard („Interpersonal Comparisons: Preference, Good, and the Instrinsic Reward of Life") und D. Davidson („Judging Interpersonal Interests").

zahl der Permutationen (ohne Wiederholung) von K. Wenn die Anzahl der anonymen gesellschaftlichen Zustände $\#Z = r$ ist und die Anzahl der Positionen jedes anonymen gesellschaftlichen Zustandes z aus Z der Anzahl der Individuen aus K entspricht, dann hat die der Menge der anonymen gesellschaftlichen Zustände Z entsprechende Menge der Arrow-Alternativen X die Mächtigkeit $rn!$.

6) *Individuelle Position.* Den Sachverhalt, daß eine Person i aus K in einer bestimmten Position p ist, kennzeichnen wir durch (i,p).

Mit Hilfe des Positionenbegriffes lassen sich unterschiedliche Arten individueller Präferenzen formal charakterisieren. Die in der Logik kollektiver Entscheidungen fast ausschließlich gebrauchten *einfachen individuellen Präferenzen* beziehen sich auf die Menge der Arrow-Alternativen. Dabei bleibt offen, welche Motivation diese Präferenzen prägt. Nun kann nicht verwehrt werden, spezifische Interpretationen dieser einfachen individuellen Präferenzen vorzunehmen. Besser wäre es jedoch, wenn man unterschiedliche Arten individueller Präferenzen auch explizit formal charakterisieren könnte. Tatsächlich wird das durch den Positionenbegriff zu einem gewissen Grad möglich.

„Empathie" ist eine Disposition (oder auch ein mentaler Zustand), die man als „Fähigkeit, sich in die Situation einer anderen Person hineinzuversetzen" umschreiben könnte. Diese Fähigkeit ist für das moralische Verhalten wesentlich. Für eigenorientierte individuelle Präferenzen spielt die Situation anderer Personen nicht direkt, sondern höchstens indirekt, nämlich als kausal wirkende Behinderung oder Förderung der eigenen Zielsetzungen eine Rolle. Der Unterschied zwischen eigenorientierten und von Sympathie geprägten individuellen Präferenzen läßt sich mit Hilfe dieser Begrifflichkeit näher charakterisieren.

Sei g die Präferenzstruktur in K. g ordnet jeder Person aus K eine individuelle Präferenzrelation über die Menge der Arrow-Alternativen X zu: $G \ni g\colon K \ni i \to g(i) \in X \times X$. Wer g kennt, weiß für jede Alternative x aus X, ob sie z. B. von der Person i gegenüber einer anderen beliebigen Alternative y aus X vorgezogen wird oder nicht. Wir nehmen an, daß die individuellen (schwachen) Präferenzen vollständig, reflexiv und transitiv sind[6].

Betrachten wir eine zweite Art individueller Präferenzen, die sich auf individuelle Positionen $(i,p) \in K \times P$ (P-Präferenzen) beziehen. Auch von diesen Präferenzen werden die Ordnungseigenschaften Reflexivität, Transitivität und Vollständigkeit gefordert. g* ist die entsprechende Präferenzstruktur, die jeder Person aus K ihre P-Präferenzrelation zuordnet. Eine individuelle P-Präferenzrelation enthält neben den Vorlieben für Positionen, die selbst eingenommen werden: $\langle (i,p),(i,q)\rangle \in g^*(i)$ auch Präferenzen über Positionen anderer Personen $\langle (j,p),(j,q)\rangle \in g^*(i)$, ja im allgemeinen

[6] Diese Begriffe wurden in § 8 eingeführt.

Fall eine Form des interpersonellen Vergleichs: $\langle(j,p),(k,q)\rangle \in g^*(i)$. „$\langle(i,p),(i,q)\rangle \in g^*(i)$" wäre etwa so zu umschreiben: „Die Person i bevorzugt für sich selbst die Position p gegenüber q".[7]

„$\langle(j,p),(j,q)\rangle \in \dot{g}^*(i)$": „i hält es für besser, wenn j in der Position p ist, als wenn j in der Position q ist".

Und der allgemeine Fall: „$\langle(j,p),(k,q)\rangle \in \dot{g}^*(i)$": „i hält es für besser, wenn j in der Position p ist, als wenn k in der Position q ist". Der interpersonelle Vergleich wird deutlicher in der Formulierung: „i hält die Position p für j besser als die Position q für k" oder „i meint, die Position p sei für j günstiger als die Position q für k".

Präferenzen dieser Art kann es natürlich auch ohne das psychologische Phänomen von Empathie und Sympathie geben, sie könnten z. B. ausschließlich auf den vermuteten Auswirkungen fremder Positionszuordnungen auf die eigene Situation beruhen. Eine Person, die an der Situation anderer Personen völlig desinteressiert ist, wäre jeweils zwischen zwei Positionszuordnungen φ und φ' bezüglich eines anonymen gesellschaftlichen Zustandes z indifferent, soweit die eigene Position unverändert bliebe: $\forall\varphi\in\Phi$: $[\varphi(i)=\varphi'(i) \rightarrow \langle(\varphi,z),(\varphi',z)\rangle \in \tilde{g}^*(i)]$. Da es eine eindeutige Zuordnung von Arrowschen Zuständen und Positionszuordnungen bei festem anonymem gesellschaftlichen Zustand z gibt, induziert g über X eine Teilrelation g_z über Φ_z – statt „$\langle\langle\varphi,z\rangle,\langle\varphi',z\rangle\rangle \in g^*(i)$" können wir daher auch kurz „$\langle\varphi,\varphi'\rangle \in g^*(i)$" schreiben. Bei gegebenem (festem) anonymen gesellschaftlichen Zustand z ergibt sich eine individuelle Präferenzrelation über die Menge der Positionszuordnungen Φ.

Die Menge der Positionszuordnungen zerfällt bei einer ausschließlich eigenorientierten Person in Äquivalenzklassen nach Maßgabe $\varphi(i)=\varphi'(i)$, und zwischen den Präferenzzuordnungen unterschiedlicher Äquivalenzklassen besteht die Präferenzrelation g: $\langle\varphi,\varphi'\rangle \in g(i) \leftrightarrow \langle\varphi(i),\varphi'(i)\rangle \in g(i)$.

Wenn alle Personen jeweils nur Präferenzen dieser Art hätten, dann wäre der Positionen-Ansatz allerdings obsolet. Wir können jedoch annehmen, daß die Äquivalenzklassenbildung bei Personen i. d. R. nicht derart einfach vor sich geht bzw. daß Personen – unabhängig von einer Veränderung ihrer individuellen Position – Präferenzen über „fremde" Präferenzzuordnungen haben. Eine Person i kann etwa eine Präferenz für $\langle j,p\rangle$ gegenüber $\langle j,p'\rangle$, also eine Präferenz dafür haben, daß eine andere Person j in einer Position p anstatt in einer Position p' ist. Ja wir wollen sogar zulassen, daß sich die Präferenzen auch auf „gemischte" individuelle Positionen beziehen, etwa: $\langle(j,p),(k,p')\rangle \in g^*(i)$. Die individuellen Präferenzrelationen über $K \times P$ kennzeichnen wir durch g^*.

[7] „g" symbolisiert die schwache Präferenz („mindestens so gut wie"), während „$\dot{g}$" die strikte Präferenz symbolisiert; analog „g^*" und „$\dot{g}^*$". „$\tilde{g}$" und „$\tilde{g}^*$" symbolisieren Indifferenz. Vgl. L. Kern/J. Nida-Rümelin, Logik kollektiver Entscheidungen, München 1994, Kap. I u. 8.

Jemand kann bestimmte Positionen für viele Personen besonders interessant und erstrebenswert halten, ohne diese selbst anzustreben: eine Rangfolge von Positionen anzugeben ist erst sinnvoll, wenn geklärt ist, wer diesen Positionen zugeordnet wird. Nur die ausschließlich eigeninteressierte Person wird Präferenzen über fremde Positionszuordnungen nur insoweit haben, als diese die eigenen Zielsetzungen direkt oder indirekt beeinflussen. Andere Personen werden Präferenzen über fremde Präferenzzuordnungen haben, die durch Sympathie oder Antipathie motiviert sind. Aber auch moralische Erwägungen könnten – etwa unter dem Gesichtspunkt der Gerechtigkeit – die Präferenzen über fremde Positionszuordnungen bestimmen.

In der Literatur werden Präferenzen, die sich auf die Menge $K \times P$ beziehen, *erweiterte Präferenzen* genannt. Diese Bezeichnung ist dann sinnvoll, wenn die üblichen einfachen Präferenzen über X als eigenorientierte Präferenzen aufgefaßt werden; denn dann wären diese Präferenzen mit einer Subrelation der erweiterten Präferenzrelation g* über $K \times P$, nämlich mit g* über $\{i\} \times P$ zu identifizieren. Wenn man dagegen die ursprünglichen Präferenzen für jede Interpretation offenhält, dann werden unterschiedliche inhaltliche Merkmale der individuellen Präferenzen formal charakterisierbar. So läßt sich z. B. erst mit Hilfe der neuen Präferenzstruktur g* Eigenorientierung individueller Präferenzen explizit machen, was im Rahmen des Arrowschen Ansatzes nicht möglich wäre.

Zwischen P-Präferenzrelation und einfacher Präferenzrelation kann durch folgende Forderung ein Zusammenhang hergestellt werden: $\langle (i,p),(i,p')\rangle \in g^*(i) :\leftrightarrow \langle p,p'\rangle \in g(i)$. Da p jedoch nur *ein* Merkmal eines Arrowschen Zustandes ist – in der volkswirtschaftlichen Diskussion entspräche p z. B. der Ausstattung der Person i mit bestimmten Gütern – darf man diese einfache Präferenzrelation nicht mit der Präferenzrelation über X verwechseln.

Für erweiterte Präferenzrelationen lautet das Pareto-Kriterium: Ein Arrow-Zustand (φ,z) ist pareto-optimal $:\leftrightarrow \neg\exists\ [(\varphi',z') \in \Phi \times Z: \langle(\varphi',z'),(\varphi,z)\rangle \in P]$, wobei P die Pareto-besser-Relation ist:

DEF.: $\langle(\varphi,z),(\varphi',z')\rangle \in P :\leftrightarrow \forall i \in K: [\langle(\varphi,z),(\varphi',z')\rangle \in g(i)] \wedge \exists j \in K: [\langle(\varphi,z),(\varphi',z')\rangle \in \dot{g}(j)]$,

bzw. bei festem anonymem gesellschaftlichem Zustand:

$$\langle\varphi,\varphi'\rangle \in P :\leftrightarrow \forall i \in K: [\langle\varphi,\varphi'\rangle \in g(i)] \wedge \exists j \in K: [\langle\varphi,\varphi'\rangle \in \dot{g}(j)]$$

Betrachten wir dagegen folgende Fassung:

$$\langle(\varphi,z),(\varphi',z')\rangle \in P' :\leftrightarrow \forall i \in K: [\langle\varphi(i),\varphi'(i)\rangle \in g^*(i)] \vee \exists j \in K: [\langle\varphi(j),\varphi'(j)\rangle \in \dot{g}^*(j)]$$

bzw. bei festem anonymem gesellschaftlichem Zustand:

$$\langle\varphi,\varphi'\rangle \in P' :\leftrightarrow \forall i \in K: [\langle\varphi(i),\varphi'(i)\rangle \in g^*(i)] \wedge \exists j \in K: [\langle\varphi(j),\varphi'(j)\rangle \in \dot{g}^*(j)].$$

Die Aussage des Pareto-Kriteriums mit dieser besser-Relation P' ist eine andere als die direkte Übersetzung des Pareto-Kriteriums in die Terminologie der Positionen. Das neue Kriterium macht ausschließlich die Präferenzen über die *eigenen* Positionen zur Grundlage der Beurteilung, während das ursprüngliche Pareto-Kriterium gegenüber unterschiedlich motivierten individuellen Präferenzen offen ist. Da es nichts darüber aussagt, aus welchen Gründen die einzelnen Personen einen Zustand präferieren, kann es durchaus sein, daß eine Person nur deshalb eine Präferenz für einen bestimmten Zustand hat, weil in diesem Zustand eine andere Person besser gestellt ist, was jedoch mit dieser neuen Fassung des Pareto-Kriteriums nicht vereinbar wäre. Allerdings entspricht die neue Fassung des Pareto-Kriteriums der gängigen Interpretation des ursprünglichen: Es macht diese jedoch explizit, d.h., andere Interpretationen, die von der alten Fassung nicht ausgeschlossen wurden, sind mit der neuen nicht vereinbar.

Erweiterte Präferenzen können auf Empathie beruhen: Man könnte sich vorstellen, daß die Präferenzen über die Positionen anderer Personen die Vermutung der präferierenden Person darüber wiedergeben, was andere Personen vorziehen würden. In einer Gesellschaft, in der alle Personen über ideale empathische Fähigkeiten verfügen, wäre die *Bedingung der schwachen Identität* erfüllt:

$$\forall i, j \in K: [\langle \varphi(i), \varphi'(i) \rangle \in \dot{g}(i) \rightarrow \langle \varphi(i), \varphi'(i) \rangle \in \dot{g}(j)].$$

Man könnte diese Bedingung auch als die Respektierung der persönlichen Vorlieben anderer Personen ansehen – diese Interpretation ist aber nicht zwingend; denn auch dann, wenn ich die persönlichen Vorlieben einer Person respektiere, kann ich aus anderen Gründen der Auffassung sein, es sei angemessen, daß diese Person nicht in eine Position kommt, die sie selbst bevorzugt.

Eine weit stärkere Bedingung verlangt, daß die Personen in K übereinstimmende erweiterte Präferenzrelationen haben:

$$\forall i, j \in K: [\langle \varphi(k), \varphi'(l) \rangle \in g(i) \rightarrow \langle \varphi(k), \varphi'(l) \rangle \in g(j)],$$
wobei $k, l \in K$.

Diese Bedingung ist zwar sehr stark, aber sie verlangt nicht, wie man vielleicht auf den ersten Blick meinen könnte, daß alle Personen die gleichen Vorlieben haben, ja sie ist sogar damit vereinbar, daß jede Person ihre Präferenzen über die Menge aller Positionen völlig frei bestimmt.

Auf der Grundlage interpersoneller Positionenvergleiche läßt sich nun ein Kriterium formulieren, das im wesentlichen eine Ausweitung des Pareto-Kriteriums darstellt: *Ein Zustand x ist relativ zu den individuellen er-*

weiterten Präferenzen einer Person i S-gerecht genau dann, wenn es keinen S-gerechteren Zustand y gibt.[8]

Sei $x = \langle \varphi, z \rangle$ und $y = \langle \varphi', z' \rangle$. Wir beschränken uns wieder auf die Relationen über Φ, so daß sich die personenrelative Relation „s-gerechter" G_i wie folgt definieren läßt:

DEF.: $\langle x,y \rangle \in G_i :\leftrightarrow \exists\, \pi \in Bij(K,K)$: $[\forall j \in K: \langle \varphi(j), \varphi'(\pi(j)) \rangle \in g(i) \wedge \exists k \in K: \langle \varphi(k), \varphi'(\pi(k)) \rangle \in \dot{g}(i)]$.

Eine Alternative x ist also dann für i S-gerechter als eine Alternative y, wenn es zu jeder Person j eine andere Person $\pi(j)$ gibt, der es in der Position $\varphi'(\pi(j))$ schlechter geht[9] als es j in der Position $\Phi(j)$ geht, und es mindestens eine Person k gibt, der es – wiederum nach der Beurteilung von i – besser geht als der Person $\pi(k)$ in der Position $\varphi'(\pi(k))$. – Oder kürzer: Ein Zustand x ist dann gerechter als ein Zustand y, wenn es zwar nicht jeder Person individuell in x besser geht als in y, es jedoch eine Permutation π über K gibt, so daß es jeder Person in x besser (oder mindestens so gut) geht, wie der ihr (durch diese Permutation) zugeordneten Person in y.

Theorem: Die Pareto-Relation der ursprünglichen Fassung ist eine Teilrelation jeder dieser individuellen Relationen G_i, wenn die schwache Identitätsbedingung erfüllt ist und die einfachen Präferenzen eigenorientiert sind.

Beweis:
(bei festem anonymem gesellschaftlichem Zustand z)

1) Annahme: $\langle x,y \rangle \in P$
2) sei o. B. d. A. $x := \langle \varphi, z \rangle$ und $y := \langle \varphi', z \rangle$
3) $\forall i \in K: \langle \varphi, \varphi' \rangle \in g(i) \wedge \exists j \in K: \langle \varphi, \varphi' \rangle \in \dot{g}(j)$, wg. (1)
4) $\forall i \in K: \langle \varphi(i), \varphi'(i) \rangle \in g(i) \wedge \exists j \in K: \langle \varphi(j), \varphi'(j) \rangle \in g(j)$, wg. Eigenorientierung
5) $\exists \pi \in Bij(K,K)$: $[\forall j \in K: \langle \varphi(j), \varphi'(\pi(j)) \rangle \in g(i) \wedge \exists k \in K: \langle \varphi(k), \varphi'(\pi(k)) \rangle \in \dot{g}(i)]$, gilt wg. (4) für $\pi = id$
6) $\langle x,y \rangle \in G_i$, q. e. d. wg. (2) und (5).

Die Relationen G_i sind im allgemeinen von Person zu Person unterschiedlich, dennoch haben sie einen ethischen Charakter, der sich unter anderem darin zeigt, daß alle Personen aufgrund der Existenzquantifikation über die Menge der Permutationen in K in gleicher Weise berücksichtigt werden. Grundlage dieser Gerechtigkeitsrelation sind jedoch die individuellen P-Präferenzen der handelnden Person, wie sie sich in g(i) manifestie-

[8] Das „S" in „S-gerechter" steht für „Suppes", denn es handelt sich hier um eine Umformulierung des „grading principle" von P. Suppes, s. FN, S. 67.

[9] Natürlich ist diese Interpretation der erweiterten Präferenz nicht zwingend. Sie identifiziert die erweiterten Präferenzen mit interpersonellen Beurteilungen der persönlichen Situation anderer Personen.

ren. Die Gerechtigkeitsrelation G_i ist nicht nur *subjektivistisch*[10], sondern auch auf die handelnde Person *relativiert*. Diese beiden Eigenschaften sind logisch unabhängig voneinander: Eine Gerechtigkeitsrelation könnte subjektivistisch sein, ohne auf die handelnde Person relativiert zu sein, wie es das Beispiel des ethischen Bayesianismus zeigt. Aber es ist natürlich ebenfalls möglich, daß eine nicht-subjektivistische Gerechtigkeitstheorie ein normatives Grundprädikat enthält, das auf die handelnde Person relativiert ist.

Da G_i eine Rangordnung von Zuständen unter einem moralischen Aspekte beinhaltet, läßt sich das Kriterium der S-Gerechtigkeit zu einer konsequentialistischen ethischen Theorie komplettieren, die allerdings eine auf die handelnde Person bezogene Abschwächung vorsieht: Die moralische Rangordnung der Zustände ist nicht interpersonell invariant.

§ 21 Prärogative

Konsequentialistische Ethiker haben sich von Anbeginn mit der Kritik an den strengen Anforderungen auseinandersetzen müssen, die ihre meist utilitaristischen Theorien an die *Unparteilichkeit* der Handelnden stellen. Diese Kritik legt es nahe, eine weitere Form der Abschwächung konsequentialistischer Ethik zu prüfen, die ein Recht des Handelnden auf eine „überproportionale" Berücksichtigung der eigenen und der Interessen nahestehender Personen beinhaltet. Ein Recht dieser Art nennen wir eine *„Prärogative"*.

Zur Vereinfachung der Argumentation nehmen wir als pars pro toto das Beispiel des hedonistischen Handlungsutilitarismus. Betrachten wir eine Person, für die das eigene Wohlergehen (und das ihr nahestehender Personen) einen größeren Stellenwert hat als das einer beliebigen anderen fremden Person. Schon die Vertreter des klassischen Utilitarismus haben sich Gedanken gemacht, unter welchen Bedingungen eine solche Person dennoch im Einklang mit dem utilitaristischen Pflichtenkriterium handeln würde[11]. Ein wesentliches Argument dafür ist, die handelnde Person könne ihre eigenen Interessen und die kausalen Wirkungen ihres eigenen Handelns auf ihr persönliches Wohlergehen besser beurteilen als jede andere Person. Die vorrangige Berücksichtigung der eigenen Interessen fördere daher das Gesamtwohlergehen in höherem Maße als das völlig

[10] Dieser Gebrauch des Prädikates „subjektivistisch" ist zwar weithin üblich, aber dennoch nicht völlig unproblematisch; denn ob eine Theorie subjektivistisch ist, ist zunächst eine Frage der Interpretation des normativen Grundprädikates und ergibt sich nicht aus der *inhaltlichen Bestimmung* für sich genommen. Es ist irreführend, allein aus der Tatsache, daß eine Theorie ihr normatives Grundprädikat (hier: „G_i-gerecht") von subjektiven Wertungen abhängig macht, auf ihren subjektivistischen Charakter zu schließen. Da dieses Problem jedoch im nachfolgenden Abschnitt aufgegriffen wird, kann es hier bei dieser vorläufigen Charakterisierung bleiben.

[11] Vgl. H. Sidgwick, *The Methods of Ethics*, London ^{7}1907, Kap. 3, § 3.

unparteiische Bemühen, allen Interessen in gleicher Weise gerecht zu werden.

Zumindest die klassische Utilitarismus-Diskussion hat dabei[12] zwei unterschiedliche Fassungen dieses Argumentes nicht unterschieden: Es kann nämlich *individuell* oder *kollektiv* verstanden werden. Individuell verstanden beruht es auf der psychologischen Annahme, daß die *Motivation*, das Gesamtwohl zu maximieren, zu singulären Handlungen führt, die das Gesamtwohl jeweils *de facto nicht* maximieren, während die Motivation, die eigenen Interessen vorrangig oder „überproportional" zu berücksichtigen, eine Maximierung des Gesamtwohls de facto nicht nur nicht ausschließt, sondern sogar erst ermöglicht. Diese psychologische Annahme ist ein Merkmal der individuellen Interpretation: Der *einzelne* Handelnde maximiert das Gesamtwohl, wenn er von der Motivation geleitet handelt, seine persönlichen Interessen bis zu einem gewissen Grad vorrangig zu berücksichtigen. Falsch wäre die Formulierung: „Der einzelne Handelnde maximiert das Gesamtwohl, wenn er seine persönlichen Interessen vorrangig berücksichtigt". De facto nämlich berücksichtigt der Handelnde seine persönlichen Interessen eben nicht „vorrangig", sofern die Funktion, die durch sein Handeln maximiert wird, das jeweilige Gesamtwohl als Wert hat.

In der kollektiven Interpretation dagegen wird nicht behauptet, daß der Einzelne de facto durch eine eher eigenorientierte Motivation das Gesamtwohl maximiert, sondern daß eine Gesellschaft, in der alle Personen je individuell *ausschließlich unparteiisch motiviert* handeln, kein maximales Gesamtwohl erreichen wird. Oder positiv formuliert: Das Gesamtwohl der Gesellschaft ist höher, wenn die Mitglieder in einem gewissen Ausmaß ihre persönlichen Interessen überproportional berücksichtigen. Diese Aussage ist logisch damit vereinbar, daß die einzelnen Mitglieder das Gesamtwohl je individuell nicht maximieren[13]. Die kollektive Fassung des Argumentes wird von der individuellen impliziert. Die Umkehrung gilt jedoch nicht.

Es ist wichtig zu sehen, daß nur die individuelle Fassung des Argumentes, nach der die überproportionale Berücksichtigung der eigenen Interessen das Gesamtwohlergehen in höherem Maße fördere als das völlig unparteiische Bemühen, allen Interessen in der gleichen Weise gerecht zu werden, eine Verteidigung des strikten Konsequentialismus darstellt. Die kollektive Fassung ist dagegen mit einem strikten Konsequentialismus – denn als solcher ist der klassische Utilitarismus überwiegend verstanden worden – unvereinbar, da damit zugestanden wird, daß es unter bestimm-

[12] Vgl. J. Bentham, *An Introduction to the Principles of Morals and Legislation*, London 1789, Kap. I u. II. Die gleiche Vermengung findet sich bei J. St. Mill, *Utilitarianism*, London 1963, Kap. II.

[13] Ein Sachverhalt, der für das Koordinationsproblem eine wesentliche Rolle spielt, wie später (in Kap. 11) noch ausgeführt wird.

ten Umständen moralisch empfehlenswert ist, eine Handlung zu wählen, deren Folgen nicht optimal sind.

Ein weiterer Aspekt dieses Argumentes ist, daß es ganz wesentlich von einer Unterscheidung Gebrauch macht, die für deontologische Ethiken typisch ist, aber von konsequentialistischen Ethiken i.d.R. abgelehnt wird: Gemeint ist die moralisch unterschiedliche Beurteilung von Handlungen und Unterlassungen[14]. J. St. Mill etwa argumentiert, daß die Gelegenheiten, zum öffentlichen Wohl beizutragen, im Alltagsleben recht selten seien, so daß nur in diesen seltenen Fällen die Motivation ausschließlich auf die Maximierung des Gesamtnutzens zu richten sei, wohingegen es in allen übrigen Fällen hinreiche, die eigenen Interessen und die einiger weniger betroffener Personen zu berücksichtigen. Diese Verteidigung des strikten Konsequentialismus ist jedoch nur dann überzeugend, wenn die Alternativen sich auf die „üblicherweise" von Handelnden in Erwägung gezogenen beschränken. Die ideale moralische Person im Sinne des strikten Konsequentialismus würde dagegen alle Handlungsoptionen unabhängig von Sitte und Gewohnheit berücksichtigen: z. B. würde sie auf dem Wege zu einer Geburtstagsfeier abwägen, ob das Geld für die Blumen wirklich optimal, also gesamtnutzenmaximierend angelegt sei, oder ob eine Spende dieses Betrages für die Welthungerhilfe nicht besser wäre. Eine ideale moralische Person im Sinne des strikten Konsequentialismus könnte angesichts dieser Alternative sogar gezwungen sein, bis zur völligen eigenen Verarmung Geld an die Welthungerhilfe zu überweisen. Es soll an dieser Stelle nicht behauptet werden, daß dies inakzeptable Konsequenzen sind und somit der strikte Konsequentialismus widerlegt ist. Vielmehr geht es hier ausschließlich um ein bestimmtes konsequentialistisches Argument zugunsten einer überproportionalen Berücksichtigung der eigenen Interessen und der Interessen nahestehender Personen: Wenn eine Person für die Konsequenzen einer unterlassenen (offenstehenden) Handlung in der gleichen Weise verantwortlich ist wie für die Konsequenzen der ausgeführten Handlung, dann läßt sich dieser Versuch der Einbettung der genannten „widerspenstigen" moralischen Intuitionen in das striktkonsequentialistische Paradigma nicht durchführen.

Eine *bereichsspezifische Prärogative,* die der handelnden Person für einen spezifischen Entscheidungsbereich die überproportionale Berücksichtigung eigener Interessen einräumt, ist nur dann im Einklang mit dem konsequentialistischen Grundmodell, wenn das Abgrenzungskriterium dieses Bereiches auf Handlungskonsequenzen Bezug nimmt. Ein Entscheidungsbereich, der aufgrund eines anderen Kriteriums individuelle Handlungen der moralischen Beurteilung entzieht, ist mit dem ethischen Konsequentialismus unvereinbar, weil dann auch solche Handlungsalternativen

[14] Vgl. R. Spaemann, „Nebenwirkungen als moralisches Problem", *Philos. Jahrbuch 82* (1975) 323–335.

ins Belieben des Handelnden gestellt wären, die möglicherweise verheerende Konsequenzen zeitigen. Dabei genügt aus methodologischen Gründen, die erst später dargestellt werden, die *kontrafaktische* Möglichkeit solcher „verheerender Konsequenzen", um die Unzulässigkeit von Abgrenzungskriterien dieser Art aufzuzeigen. Ein Kriterium, das bestimmte Entscheidungsbereiche aus der moralischen Beurteilung ausgrenzt, führt zu einer gespaltenen Lebenswelt: In der einen Welt dürfen die eigenen Projekte und Zielsetzungen eine wesentliche Rolle spielen, in der anderen tritt die handelnde Person – bildlich gesprochen – aus sich heraus und verwandelt sich in einen unparteiischen Maximierer des Gesamtwohls. Eine Theorie, die das Verfolgen persönlicher Projekte *vorschreibt*, wäre daher unbefriedigend. Trotzdem sollte jede Theorie das Verfolgen persönlicher Projekte in einem bestimmten Ausmaß *erlauben*. Die interessante Frage lautet: Ist eine konsequentialistische Theorie dazu in der Lage?

Betrachten wir folgenden (sprachlich etwas umständlichen, aber mathematisch sehr einfachen) Vorschlag: Eine Person *darf* (sie ist nicht verpflichtet) eine die Aggregationsfunktion nicht maximierende Handlungsalternative wählen, wenn die Verminderung des Aggregationswertes kleiner ist als das x-fache des eigenen Vorteils im Vergleich zu der Handlungsalternative, die die Aggregationsfunktion maximieren würde[15] (wobei x keine natürliche Zahl sein muß, auch bescheidene 1,2 oder sogar 0,3 wären zulässig; denn der eigene Vorteil geht in die Summe ohnehin ein). Wenn x gegen Null strebt, dann nähert sich diese abgeschwächte Form des Konsequentialismus der strikten Form an.

Es besteht eine gewisse Verwandtschaft mit dem Gerechtigkeitskonsequentialismus: Die Berücksichtigung einer Prärogative führt – ähnlich der Berücksichtigung von Gerechtigkeitsaspekten – im Rahmen des konsequentialistischen Ansatzes zu einer Wertfunktion, die sich nicht einfach aus der Aufsummierung des auf die einzelnen Personen verteilten außermoralischen Guten ergibt, sondern in die eine besondere und ungleiche Berücksichtigung einzelner Personen eingeht. Aber es besteht zugleich ein wichtiger und charakteristischer Unterschied zwischen diesen beiden Formen des Konsequentialismus: Wie oben aufgezeigt, gehört der Gerechtigkeitskonsequentialismus nicht zu den teleologischen Theorien, weil diese Modifikation der Wertfunktion unter einem *moralischen Aspekt* erfolgt und damit das „Gute" nicht mehr unabhängig vom „Rechten" ist, während die Modifikation der Wertfunktion durch die Berücksichtigung einer Prärogative nicht aufgrund eines moralischen Kriteriums erfolgt, sondern

[15] Einen Vorschlag dieser Art unterbreitet S. Scheffler in: *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*, Oxford 1982, S. 20ff. Der Titel ist insofern irreführend, als Scheffler eine Theorie vertritt, die Handeln im (strikt-)konsequentialistischen Sinne *immer erlaubt*, aber nicht immer vorschreibt (vgl. S. 80–114). Von einer Widerlegung des Konsequentialismus erwartet man jedoch den Nachweis, daß es *nicht immer erlaubt ist, im Sinne des Konsequentialismus* zu handeln.

eher „pragmatische" oder anthropologische Gründe hat: Einer Person kann Selbstlosigkeit in dem Ausmaß, wie sie von Theorien des strikten Konsequentialismus gefordert wird, nicht zugemutet werden.

8. Kapitel: Folgerungen für die Typologie ethischer Theorien

Aus der vorgenommenen Charakterisierung des ethischen Konsequentialismus ergeben sich Folgerungen für die Klassifizierung ethischer Theorien generell, die hier – um Mißverständnisse auszuschließen – erwähnt werden müssen.

§ 22 *Methodischer Status*

Theorien der Ethik lassen sich nach ihrem methodischen Status in drei Gruppen einteilen. Eine wichtige Tradition ethischen Denkens beschränkt sich auf die *Erklärung moralischer Phänomene,* insbesondere auf die des moralischen Handelns. Wir nennen solche Theorien *explanatorisch.* Bisweilen werden Ethiken dieser Art auch als naturalistisch bezeichnet. Dies ist deswegen mißverständlich, weil man im Sinne der Theorie des naturalistischen Fehlschlusses von naturalistischen Theorien auch in einem ganz anderen Kontext spricht: Naturalistische Theorien wären demnach solche Theorien, die vorgeben, normative Urteile (explizit oder implizit) allein aus deskriptiven Urteilen logisch ableiten zu können[1]. Die logische Trennung von normativen und deskriptiven Urteilen geht auf David Hume zurück, dessen Ethik sich auf die *Erklärung moralischer Phänomene,* insbesondere auf die moralischen Handelns beschränkt. Man kann David Hume daher als ersten Vertreter einer explanatorischen Ethik bezeichnen[2]. Vor Hume war die methodische Trennung explanatorischer und normativer Urteile nicht üblich. Die schottische Moralphilosophie ist dafür ein besonders eindrucksvoller Beleg[3]. Der Übergang von einer explanatorischen (philosophischen) Ethik zu einer soziologischen (psychologischen oder ethnologischen) Ethik ist nicht trennscharf. Soziologische Untersuchungen haben konkrete gesellschaftliche Normensysteme in einem bestimmten historischen Kontext zum Gegenstand, während sich explanatorische Ethik mit der Entwicklung eines generell anwendbaren Erklärungsmusters moralischer Phänomene befaßt. Das Verhältnis explanatorische Ethik vs. soziologische Ethik entspricht in etwa dem Verhältnis philosophischer vs. biologischer Anthropologie. Man kann soziologische (und psychologische) sowie explanatorische Ethik zusammen als *empirische Ethik* bezeichnen. Trotz dieser Charakterisierung wäre es jedoch nicht angemessen, diese Disziplin in toto aus dem Bereich der Philosophie auszuschließen. Auch konsequentialistische Ethiken können einen explanatorischen Status

[1] Vgl. G. E. Moore, *Principia Ethica,* Cambridge 1903, Kap. 2.
[2] Vgl. D. Hume, *An Enquiry Concerning the Principles of Morals,* London 1751.
[3] Vgl. Shaftesbury, *An Inquiry concerning Virtue or Merit,* London 1711; F. Hutcheson, *An Inquiry into the Originals of Our Ideas of Beauty and Merit,* London 1925; J. Butler, *Fifteen Sermons,* London 1926.

haben. In diesem Fall erklären sie moralisches Verhalten als erwartungswertmaximierend im Hinblick auf eine bestimmte subjektive Wertfunktion, die natürlich erst noch genauer als utilitaristisch, egoistisch, hedonistisch etc. zu spezifizieren ist.

Normativ-axiologische Theorien beanspruchen dagegen, ein System der moralischen Beurteilung zu entdecken (im Falle kognitivistischer Theorien[4]) oder zu erfinden (im Falle non-kognitivistischer Theorien). Uns interessiert der konsequentialistische Ansatz in der Ethik hier ausschließlich unter dem Gesichtspunkt der moralischen Beurteilung. Zwischen der explanatorischen und der normativ-exiologischen Ebene der Analyse gibt es allerdings Verbindungen: Eine gute normativ-axiologische Theorie der Ethik hat i. allg. auch den Anspruch, explanatorisch relevant zu sein[5] – umgekehrt spricht es gegen den normativ-axiologischen Gehalt einer Theorie, wenn sie keinerlei explanatorische Funktion hat. Eine ethische Theorie, die beansprucht, normativ-axiologisch adäquat zu sein, sollte auch zu einer Erklärung von Phänomenen wie den folgenden beitragen: der spezifischen Verwendung moralischer Ausdrücke; der Rolle gesellschaftlicher Kooperation; der Funktion moralischer Institutionen; der Neigung, unter einem moralischen Gesichtspunkt auf Interessen anderer Personen Rücksicht zu nehmen; oder der sittlichen Beurteilung von Charaktereigenschaften. Im nachfolgenden Teil III werden einige dieser Aspekte einer genaueren Analyse unterzogen.

Darüber hinaus gibt es Theorien, die sich ausschließlich mit der Bedeutung moralischer Ausdrücke und der logischen Analyse ethischer Begründungen auseinandersetzen[6]. Soweit sich diese Theorien ihrerseits auf ethische Theorien beziehen, haben sie den Status einer *Metatheorie:* So wie sich die Wissenschaftstheorie der Physik mit der logischen Struktur und der Genese sowie den Begründungsverfahren physikalischer Theorien auseinandersetzt, so steht auch für *metaethische Theorien* die *rationale Rekonstruktion* und die *Explikation der Grundbegriffe* ethischer Theorien im Mittelpunkt. Unter Metaethik wird jedoch auch die – meist sprachphilosophische – *Analyse moralischer Ausdrücke* der Alltagssprache verstanden[7]. Im 13. Kapitel befassen wir uns dagegen mit einem besonderen Typus der ethischen Analyse, den wir als *Kritik zweiter Ordnung* bezeichnen. Eine ethische Kritik zweiter Ordnung unterscheidet sich von einer metaethischen Analyse darin, daß sie keinen primär deskriptiven, sondern einen *normativen Status* hat.

[4] Ein interessantes Plädoyer für eine kognitivistische Ethik enthält F. v. Kutschera, *Grundlagen der Ethik*, Berlin 1982.

[5] Besonders deutlich wird dies bei J. Rawls, *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975, Kap. 8.

[6] So z. B. P.-H. Nowell-Smith, *Ethics*, Harmondsworth 1954.

[7] Als exponiertester Vertreter dieser Art ethischer Analyse kann nach wie vor R. M. Hare, *The Language of Morals*, Oxford 1952 und *Freedom and Reason*, Oxford 1963 gelten.

§ 23 Subjektiv und objektiv

Man spricht bisweilen davon, eine bestimmte Handlung sei subjektiv richtig, jedoch objektiv falsch gewesen. Unter „subjektivistische Ethik" im engeren Sinne kann man solche Theorien fassen, die das, was moralisch richtig ist, grundsätzlich auf die Ziele der handelnden Person relativieren. Im weiteren Sinne aber kann man auch jede Theorie subjektivistisch nennen, die ihre normativen oder axiologischen Prädikate so faßt, daß die normativen (axiologischen) Urteile bzw. Behauptungen der Theorie sich aus Sätzen über subjektive Ziele und Präferenzen (nicht nur der handelnden Person) logisch ableiten lassen[8]. Bei genauerer Betrachtung stellt sich allerdings heraus, daß dieser Sprachgebrauch nicht eindeutig ist. In der Bayes'schen Entscheidungstheorie wird der Rationalitätsbegriff auf die subjektiven Wahrscheinlichkeiten der handelnden Person relativiert, d.h., rational ist diejenige Handlung, die bei den gegebenen subjektiven Präferenzen der Person und ihren subjektiven Wahrscheinlichkeiten einen maximalen Erwartungswert ergibt. Dies ist offensichtlich nicht im Einklang mit vortheoretischen Rationalitätsbedingungen wie z. B. der, daß eine rationale Person bemüht sein sollte, sich ein „richtiges Bild" von den Wahrscheinlichkeiten der relevanten Umstände zu machen. Dieses „richtige Bild" ist aber nicht immer schon dann gegeben, wenn die subjektiven Wahrscheinlichkeiten der handelnden Person mit den Kolmogorov-Axiomen der Wahrscheinlichkeitstheorie übereinstimmen.

Betrachten wir folgendes Beispiel einer moralisch relevanten Entscheidungssituation: Hans steht 1944 auf einer Brücke und sieht, wie jemand von der Strömung mitgerissen wird und zu ertrinken droht. Hans, ein sehr guter Schwimmer, zögert nicht, sondern springt in den Fluß und rettet die Person. Hans hat sicherlich moralisch richtig gehandelt. Nach der Rettung stellt sich heraus, daß es sich bei der geretteten Person um Adolf Hitler handelt. Nun kann man die Auffassung vertreten, daß das Ertrinken Hitlers zu diesem Zeitpunkt unter moralischen Gesichtspunkten seiner Rettung vorzuziehen ist. Wenn diese Auffassung, wie sie z. B. von den Verschwörern des 20. Juli 1944 um Graf Stauffenberg vertreten wurde, zutrifft, dann ist es objektiv falsch, Hitler zu retten. Dennoch wird man Hans für seine Rettungsaktion keinen moralischen Vorwurf machen. Da er auf der Brücke nicht ahnen konnte, daß es sich bei dem Ertrinkenden um Hitler handelte, war seine Handlung subjektiv richtig. Eine ethische Theorie hat offensichtlich beide Aufgaben: Kriterien dafür zu entwickeln, was objektiv$_1$ richtig ist, und Kriterien dafür zu entwickeln, was subjektiv$_1$

[8] *Ethica ordine geometrico demonstrata* (1677), Teil IV u. V von Baruch de Spinoza ist ein besonders eindrucksvolles Beispiel einer subjektivistischen Ethik, die aber in ihrer Zuordnung nicht eindeutig ist, da Spinoza seine ethischen Postulate sowohl aus den Interessen und subjektiven Vorstellungen der handelnden Person als auch des Menschen als solchem ableitet. Vernünftige Handlungen sind gut, weil es für den Menschen am besten ist, vernünftig zu handeln.

richtig ist („objektiv$_1$" und „subjektiv$_1$" verwendet wie in diesem Beispiel).

Wir modifizieren das Beispiel: Hans sieht auf der Brücke, daß es sich um Hitler handelt. Hans, ein glühender Verehrer des „Führers", ist gern bereit, sein Leben für dessen Rettung aufs Spiel zu setzen, ohne dabei auf eine spätere Gratifikation seines Mutes zu hoffen. Ungeachtet seiner persönlichen Interessen hält es Hans vielmehr für seine Pflicht, den Führer zu retten. Hans handelt richtig im Sinne seiner subjektiven Moral. Man kann auch sagen, Hans handelt subjektiv richtig. Dies ist aber eine ganz andere Verwendung von subjektiv als im obigen Beispiel. Diese zweite Art subjektiv richtigen Handelns ist offensichtlich i. allg. nicht Gegenstand einer Theorie normativ-axiologischer Ethik, sondern höchstens Gegenstand explanatorischer – oder weil es sich hier um einen sehr speziellen Fall handelt, der die Berücksichtigung der konkreten Zeitumstände und gesellschaftlichen Bedingungen ebenso verlangt, wie eine genauere Kenntnis der psychischen Merkmale von Hans – einer eher soziologischen oder psychologischen Ethik. Man kann jedenfalls von einer normativ-axiologischen Ethik nicht erwarten, daß sie Kriterien dafür entwickelt, was jeweils subjektiv, moralisch richtig sein könnte. Subjektiv$_2$ richtig heißt hier: richtig im Sinne seiner subjektiven Moral.

Modifizieren wir das Beispiel erneut: Hans ist kein besonders guter Schwimmer und die Stromschnellen sind reißend. Dennoch springt er ins Wasser und kommt zusammen mit der ertrinkenden Person um, ohne gewußt zu haben, um wen es sich handelt. Im Gegensatz zur ersten Situation kann man jetzt sagen, Hans hätte wissen müssen, daß seine Handlungsweise unvernünftig ist. Sein Fehler war, daß er die Situation völlig falsch eingeschätzt hat. Trotzdem hat er subjektiv$_3$ richtig gehandelt – er wollte den Ertrinkenden retten und glaubte (fälschlich), ihn retten zu können.

Während man im Falle des „subjektiv$_1$ moralisch richtigen" Handelns von Hans interessanterweise auch sagen kann, er habe *objektiv* richtig gehandelt, da er nicht wissen konnte, daß es sich um Hitler handelt, ist dies bei den beiden anderen Varianten – „subjektiv$_2$ moralisch richtig" und „subjektiv$_3$ moralisch richtig" – sprachlich nicht möglich. Um diesen tiefen Unterschied deutlich zu machen, ist es zweckmäßiger, von „objektiv im starken Sinne" (zuvor „objektiv") bzw. von „objektiv im schwachen Sinne" (zuvor „subjektiv$_1$") und von „subjektiv im starken Sinne" (zuvor „subjektiv$_2$") bzw. von „subjektiv im schwachen Sinne" (zuvor „subjektiv$_3$") zu sprechen.

Wenn der Handelnde über vollständiges Wissen verfügt, d. h. wenn dem Handelnden alle für die moralische Beurteilung relevanten Merkmale der Situation bekannt sind, dann impliziert moralisch richtiges Handeln im stark-objektiven Sinn auch moralisch richtiges Handeln im schwach-objektiven Sinn. Ohne vollständige Information ist es jedoch möglich, daß er im starken Sinne objektiv richtig handelt, obwohl er im

schwachen Sinne nicht richtig handelt. Man betrachte etwa folgendes Beispiel: Wenn ein Einbrecher durch seinen Einbruch zufällig verhindert, daß der Ehemann seine Frau im Affekt tötet, dann ist die Verhinderung dieses Totschlags objektiv richtig. Dennoch ist der Einbruch nicht objektiv richtig im schwachen Sinne.

Das oben genannte Beispiel könnte das Mißverständnis nahelegen, daß der Unterschied zwischen objektiv im starken Sinne und objektiv im schwachen Sinne nur im Rahmen konsequentialistischer Ethik-Konzeptionen Sinn macht. Ohne Zweifel ist der Unterschied bei konsequentialistischen Konzeptionen leichter zu charakterisieren: Objektiv im starken Sinne bedeutet im Rahmen einer konsequentialistischen Theorie, daß eine Handlungsweise objektiv die (moralisch) besten Konsequenzen hat, während objektiv in schwachem Sinne schon dann gegeben ist, wenn der Handelnde diejenige Handlung wählt, die angesichts der gegebenen Entscheidungssituation die besten Konsequenzen *erwarten* läßt. Es ist jedoch auch im Rahmen nicht-konsequentialistischer Theorien möglich, stark-objektive Kriterien zu entwickeln. Die Verallgemeinerungsform des Regelutilitarismus (entscheide dich für diejenige Handlung, deren allgemeine Ausführung die besten Konsequenzen hat) formuliert z. B. ein stark-objektives Kriterium. Der von Kant entwickelte Typus einer deonotologischen Ethik beschränkt sich dagegen ausschließlich auf schwach-objektive Kriterien. Eine Pflichtenethik wird sich immer auf schwach-objektive Kriterien beschränken müssen, nicht jede deontologische Ethik ist jedoch als Pflichtenethik zu charakterisieren.

Das Verhältnis von objektiv im starken und im schwachen Sinne ist davon abhängig, ob man den grundlegenden Naturgesetzen einen deterministischen oder probabilistischen Charakter zumißt. Bei einer deterministischen Auffassung ist die Kluft zwischen Objektivität im starken und im schwachen Sinne deutlich größer: Eine Handlung h, deren Folgen nicht optimal sind, ist schon zum Zeitpunkt ihres Vollzugs objektiv im starken Sinne falsch; denn die moralische Person würde bei vollständiger Information anders handeln. Bei einer probabilistischen Auffassung hingegen könnte es sein, daß die Person h aufgrund richtiger Wahrscheinlichkeitseinschätzungen vollzieht und h den (moralischen) Erwartungswert maximiert. Die nachträgliche Information, daß die Folgen der Handlung schlecht sind, besagt bei einer probabilistischen Auffassung daher nicht, daß die Person aufgrund falscher Annahmen entschieden hat. Im Rahmen einer probabilistischen Auffassung verschiebt sich die Bedeutung des Prädikates „schwach-objektiv" in Richtung starker Objektivität.

Selbst für eine deterministische Auffassung sind stark-objektive Kriterien nur eine Vorstufe der ethischen Theoriebildung. Der Grund dafür ist folgender: Jede ethische Theorie T ist wesentlich durch die durch sie definierte *ideale moralische Person* charakterisiert. Eine normative-axiologische Theorie, die nicht angeben kann, was eine Person auszeichnet, die

mit T übereinstimmt, kann nicht als Theorie normativer Ethik gelten. Insofern ist es ein Adäquanzkriterium jeder Theorie T, daß nicht der bloße Zufall über den Grad der Übereinstimmung eines Menschen mit der durch T definierten idealen moralischen Person entscheidet. Aufgrund der Begrenztheit menschlichen Wissens ist aber die Übereinstimmung mit stark objektiven Kriterien auch im Rahmen einer deterministischen Auffassung in hohem Maße vom Zufall abhängig. Diese Überlegung stuft stark-objektive Kriterien als zum Bereich vornormativer Betrachtungen gehörend ab. Kriterien stark- und schwach-subjektiver moralischer Richtigkeit sind dagegen ebenfalls nur am Rande Gegenstand normativ-axiologischer Ethik – sie gehören eher zu den Bereichen der Lerntheorie und der empirischen Ethik, so daß *schwach-objektive Kriterien im Zentrum der ethischen Analyse stehen.*

§ 24 Gegenstände

Üblicherweise werden ethische Theorien *ihrem Gegenstand entsprechend* in drei Gruppen eingeteilt: *Handlungs-, Motivations-* und *Tugendethiken.* Der Gegenstand der normativen Kriterien von Theorien der ersten Gruppe sind Handlungen, der der zweiten (Handlungs-) Motive und der der dritten Charaktermerkmale und Verhaltensdispositionen. Man könnte als eine vierte Gruppe diejenigen Theorien hinzufügen, deren Gegenstand *Institutionen* sind. Insoweit eine ethische Theorie nicht ausschließlich theoretischen Charakter hat, enthält sie Verpflichtungsurteile, d. h. Urteile darüber, was getan werden sollte. Das System von Verpflichtungsurteilen einer Theorie T ist der *normative* Gehalt von T. Theorien der Ethik, die weder zur empirischen Ethik noch zur Metaethik gehören, aber dennoch explizit keine Verpflichtungsurteile enthalten oder implizieren, sind nicht normativ, sondern axiologisch: D. h., sie formulieren moralische Werturteile.

Konsequentialistische Theorien sind zunächst offen gegenüber dem Gegenstand der moralischen Beurteilung: Man kann Institutionen unter dem Aspekt der Optimierung der Folgen moralisch beurteilen, ebenso wie Einzelhandlungen, Gebräuche, Sitten, Handlungsregeln, Dispositionen, Tugenden, Charaktere, Persönlichkeitsmerkmale, Institutionen etc. Die moralische Beurteilung dieser Gegenstände dient im Rahmen einer Theorie normativer Ehtik dazu, Kriterien für die moralische Beurteilung freien Handelns zu entwickeln. Es ist nicht der Gegenstand der moralischen Beurteilung, der eine Theorie aus dem Bereich der konsequentialistischen Theorien ausschließt, sondern es ist die Art und Weise der Übertragung der moralischen Beurteilung auf den Gegenstand freien Handelns. Konsequentialistisch sind nur solche Theorien, die das freie Handeln selbst, genauer: die jeweils in Frage stehende singuläre Handlung *unter dem Aspekt der Folgenoptimierung* moralisch beurteilen. Wenn Abgeordnete z. B. vor

der Alternative stehen, die eine oder die andere Gesetzgebungsinitiative zu starten, dann können die beiden geplanten Gesetze unter dem Aspekt der Folgen konsequentialistisch bewertet werden. Diejenige Handlung ist dann konsequentialistisch geboten, die zum Gesetz mit den optimalen Folgen führt. Dieses Gesetz selbst wird mit Sanktionen verbunden sein, also i. d. R. bestimmte Verhaltensweisen in der Gesellschaft fördern und andere Verhaltensweisen zurückdrängen oder verhindern. Hier hat die moralische Beurteilung bestimmter Verhaltensweisen Auswirkungen auf die moralische Beurteilung einer singulären Handlung (welche Gesetzgebungsinitiative gewählt wird). Aber dies geschieht unter dem Aspekt der kausalen Folgen einer singulären Handlung für die Ausprägung bestimmter Verhaltensweisen und bleibt insofern im konsequentialistischen Kontext. Eine Übertragung ganz anderen Charakters nimmt dagegen der Regelutilitarismus sowie die Theorie der utilitaristischen Verallgemeinerung vor: Die utilitaristische Beurteilung von Regeln überträgt sich hier zwar ebenfalls auf die moralische Beurteilung singulärer Handlungen, aber diese Übertragung ist nicht von konsequentialistischem Typus. Der Grund liegt einfach darin, daß die moralische Beurteilung der singulären Handlung sich weder im Regelutilitarismus noch in der Theorie der utilitaristischen Verallgemeinerung am Kriterium der Folgenoptimierung orientiert: Es geht nicht um die Etablierung von Regeln als Folge einer singulären Handlung, vielmehr gelten singuläre Handlungen nur dann als moralisch erlaubt, wenn sie im Einklang mit einer konsequentialistisch optimalen Regel stehen. Die Handlung selbst ist damit aber noch nicht konsequentialistisch optimal, und insofern ist der Regelutilitarismus keine konsequentialistische Theorie. Daher gibt es unter dem Aspekt des normativen Kriteriums (des Kriteriums der Verpflichtung) eine enge Verwandtschaft zwischen dem Kantischen Typ der Moraltheorie und dem Regelutilitarismus, während sich die Gemeinsamkeit der beiden Utilitarismus-Versionen auf den axiologischen Teil, nämlich die Theorie des außermoralisch Wertvollen beschränkt.

Es ist nicht sinnvoll, Verpflichtungsurteile auf den „äußeren Aspekt" von Handlungen zu beschränken. Es macht einen Unterschied aus, ob ich meinen Arm hebe oder ob sich mein Arm hebt, etwa als Folge eines affizierten elektromagnetischen Impulses oder eines unwillkürlichen Muskelkrampfes. Wenn ich meinen Arm hebe, entspricht dem eine Handlung, andernfalls geschieht mir etwas – aber was da geschieht, kann mir nicht als Handlung angerechnet werden. Aber auch dann, wenn man davon ausgehen kann, daß es sich nicht um eine unwillkürliche Körperbewegung, sondern um absichtliches$_0$ Verhalten handelt, erfolgt die Identifikation bestimmter Verhaltensvorkommnisse mit dem äußeren Aspekt einer Handlung aufgrund vermuteter intentionaler mentaler Zustände. Motivierende und vorausgehende Absichten gehören zur vollständigen empirischen Erfassung menschlichen Verhaltens, das in seiner ganzen Komple-

xität Gegenstand des ethischen Verpflichtungsurteils sein muß. Auch konsequentialistische Normen beziehen sich, wie wir gesehen haben, auf ganze Komplexe motivierender und vorausgehender Absichten. Eine adäquate ethische Beurteilung läßt sich nicht auf den „äußeren" – d. h. kausal wirksamen – Aspekt des Handelns beschränken. Die Trennung von Handlungs- und Motivethiken ist daher systematisch unergiebig. Allerdings muß zwischen einer Einteilung nach Gegenstandsarten und einer Einteilung nach dem *Typus normativer Kriterien* streng unterschieden werden. Wenn eine Theorie Handlungen ausschließlich nach ihrer motivierenden Absicht beurteilt, dann sind nicht Motive *primärer Gegenstand* dieser Theorie, vielmehr werden *Handlungen primär aufgrund ihrer Motive beurteilt*. Wenn von „Motiv-" oder „Tugendethik" die Rede ist, muß jeweils geprüft werden, ob damit der Typus von Verpflichtungskriterien oder der Gegenstand der Theorie gemeint ist[9].

Konsequentialistische Theorien sind nach unserem Verständnis – soweit sie keinen explanatorischen Status haben – immer normativ. Ihre Verpflichtungskriterien ergeben sich allerdings in fast trivialer Weise in Gestalt der Erwartungswert-Maximierung aus dem axiologischen Teil der Theorie: Es gibt keine rein axiologischen konsequentialistischen Ethiken. Theorien, die ausschließlich andere Gegenstände als Handlungen einer moralischen Beurteilung unterziehen und dabei in der einen oder anderen Weise die kausalen Folgen und ihren (außermoralischen) Wert zugrunde legen, sollte man nicht „konsequentialistisch" nennen. Das Unterscheidungsmerkmal konsequentialistischer Theorien bezieht sich ausschließlich auf den Typus des Verpflichtungskriteriums. Das konsequentialistische Verpflichtungskriterium wird auf konkrete Einzelhandlungen angewendet, nicht auf Handlungsweisen; denn die vorausgehenden Absichten werden durch eine absichtliche$_o$ Handlung erfüllt – absichtlich$_o$ ist aber der konkrete Vorgang der Handlung, nicht die Handlungsweise.

§ 25 *Teleologisch, deontologisch und konsequentialistisch*

Die wichtigste Modifikation, die sich aus unserer Charakterisierung der konsequentialistischen Ethik für die Typologie ethischer Theorien ergibt, betrifft die Einteilung teleologische vs. deontologische Theorien. Es wurde gezeigt, daß sich in dieser Gegenüberstellung verschiedene Unterscheidungskriterien verbergen, die nicht konfundiert werden dürfen. Alle Kriterien beziehen sich natürlich auf Theorien normativer Ethik (auf ethische Theorien, die keine Verpflichtungskriterien enthalten, lassen sich diese Prädikate nicht anwenden).

[9] Der von R. M. Adams entwickelte „Motive Utilitarismus", *J. Phil* 73 (1976) 467–481 unterscheidet sich z. B. vom klassischen Utilitarismus nicht im Gegenstand der moralischen Beurteilung, sondern ausschließlich im Typus des Verpflichtungskriteriums.

Ein Unterscheidungskriterium bezieht sich darauf, *ob für eine normative Theorie T das moralisch Gute abhängig vom Rechten ist* bzw. ob die moralisch-axiologischen Aussagen von T abhängig von den normativen sind. Wird diese Frage bejaht, dann handelt es sich um eine *deontologische* Theorie. Nicht jede normative Theorie, die nicht deontologisch ist, ist zugleich teleologisch; denn sonst wäre eine normative Theorie, die z. B. ein Pflichtenkriterium enthält und eine davon unabhängige Theorie des moralischen Wertes von Charaktermerkmalen, teleologisch. Das Prädikat „teleologisch" sollte jedoch für solche Theorien normativer Ehtik reserviert bleiben, für die erstens das Gute außermoralisch verstanden werden kann, und zweitens das Rechte der Maximierung des (außermoralisch) Guten dient.

Eine Theorie normativer Ethik ist teleologisch genau dann, wenn das Rechte das (außermoralisch) Gute maximiert.

Es kann deontologische Theorien geben, die zugleich konsequentialistisch sind (s. § 19). Die Prädikate „teleologisch" und „deontologisch" markieren zudem keine vollständige Dichotomie. Zwar gibt es keine Theorie, die zugleich teleologisch und deontologisch ist, aber es gibt Theorien, die weder teleologisch noch deontologisch sind (z. B. Tugendethiken).

Teleologische Theorien sind ein Spezialfall konsequentialistischer Theorien. Das Prädikat „konsequentialistisch" zeichnet solche Theorien aus, für die sich der moralische Wert einer Handlung ausschließlich nach dem Wert ihrer Folgen bemißt, und zwar unabhängig davon, ob diese Werte einen außermoralischen (wie bei teleologischen Theorien) oder einen moralischen Status haben, aber auch unabhängig davon, ob die Theorie der moralischen Werte von der des Rechten abhängt (wie bei deontologischen Theorien).

Jede teleologische Theorie ist auch konsequentialistisch, aber nicht jede konsequentialistische Theorie ist teleologisch.

Eine Theorie normativer Ethik ist konsequentialistisch genau dann, wenn das Rechte das Gute maximiert.

Teil III:
Kritik des Konsequentialismus in der Ethik

9. Kapitel: Integrität der Person

Die Persönlichkeit eines erwachsenen Menschen ist im allgemeinen durch dauerhafte Bindungen an Personen und Institutionen, durch persönliche Zielsetzungen und längerfristige Projekte bestimmt. Eine ethische Theorie, die nur von Personen erfüllt werden kann, die keine persönlichen Bindungen und Projekte kennen, verletzt die *Integrität der Person* und ist daher aus anthropologischen, praktischen und theoretischen Gründen inadäquat: Aus anthropologischen, weil persönliche Bindungen und Projekte ein wesentliches Merkmal eines reifen Menschen sind; aus praktischen, weil die Verpflichtungskriterien einer solchen Theorie keine Chance haben, befolgt zu werden; und aus theoretischen, weil die Verwirklichung einer guten Lebensform (eines gelungenen Lebens) zu den Intentionen moralischen Handelns zählt[1].

Der strikte ethische Konsequentialismus kann als Paradigma einer Theorie gelten, die mit der so verstandenen Integrität der Person nicht vereinbar ist. Dieser Einwand wurde schon – in unterschiedlichen Varianten und unter anderen Etikettierungen – gegen den klassischen Utilitarismus vorgebracht. In den letzten Jahren hat B. Williams[2] dieses Argument verschärft und ausdrücklich nicht nur gegen den Utilitarismus, sondern

[1] Ein gutes Leben ist für die traditionelle teleologische Ethik die primäre Intention moralischen Handelns. Dies ist mit einer pluralen Werttheorie durchaus vereinbar: Für unterschiedliche Personengruppen sind unterschiedliche Lebensformen angemessen, vgl. Platon, *Politeia*, bes. 4. Buch und, weniger deutlich, Aristoteles, *Politik*, 7. Buch. Moderne konsequentialistische Ethiken vertreten dagegen zumeist eine subjekt-relative Werttheorie, besonders deutlich in der Theorie des Präferenzutilitarismus von J. C. Harsanyi, *Essays on Ethics, Social Behavior and Scientific Explanation*, Dordrecht 1978. Aber auch deontologische Theorien enthalten explizit oder implizit eine Theorie des moralisch (und u. U. auch außermoralisch) Wertvollen. Die Pflichten deontologischer Normensysteme dienen i. allg. auch der Verwirklichung des Guten. Das gilt z. B. für die deontologischen Theorien von D. Ross, *The Right and the Good*, Oxford 1930 oder J. Rawls, *A Theory of Justice*, Oxford 1972; dt. *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975.

[2] Die Utilitarismus-Kritik B. Williams' beruht fast ausschließlich auf diesem Argument, vgl. B. Williams, *Kritik des Utilitarismus*, Frankfurt/M. 1979 (Orig. Smart/Williams 1973). Williams beantwortet einige Einwände gegen seine Darstellung in: *Moralischer Zufall*, Königstein/Ts. 1984, Kap. 3 (Utilitarismus und moralische Selbstgefälligkeit, Orig. in: *Contemporary British Philosophy*, hg. von H. D. Lewis, London 1976). Zur Ausweitung der Kritik (allerdings ohne Bezugnahme auf den Begriff der Integrität) vgl. B. Williams, „Persons, Character, and Morality", in: *The Identity of Persons*, hg. v. A. Rorty, Berkeley u. a. 1976. Einen ähnlichen Standpunkt wie Williams vertritt M. Stocker, „The Schizophrenia of Modern Ethical Theories", *Journal of Philosophy* 73 (1976) 453–466; in abgeschwächter Form L. Lomasky, „A Refutation of Utilitarianism, *Journal of Value Inquiry* 17 (1983) 259–281 und W. Wilcox, „Egoists, Consequentialists, and Their Friends", *Philosophy and public Affairs* 16 (1987) 73–84. Zur Kritik dieser These aus utilitaristischer Sicht vgl. etwa D. Brink, „Utilitarian Morality and the Personal Point of View", *Journal of Philosophy* 83 (1986) 417–438.

gegen konsequentialistische Ethiken generell gewendet. Williams Bemerkungen zu diesem Problem und insbesondere seine Beispielfälle haben eine große Suggestivkraft, die allerdings verdeckt, daß er keine systematische Theorie dieser Problematik entwickelt[3]. Williams konzentriert seine Kritik auf Beispiele, die eines gemeinsam haben, „daß nämlich dann, wenn der Handelnde nicht etwas Widerwärtiges tut, jemand anders es tun wird"[4]. Wir unterscheiden dagegen zwei Aspekte der Integritätsproblematik konsequentialistischer Theorien. Der erste Aspekt betrifft den strikten Konsequentialismus in der Ethik, während der zweite den Konsequentialismus insgesamt in Frage stellt. Für beide Aspekte spielt die dem Konsequentialismus innewohnende extreme und inadäquate Theorie der Verantwortlichkeit eine Rolle.

§ 26 *Überbordende Verantwortlichkeit*

Die ideale moralische Person einer konsequentialistischen Theorie wählt diejenige Handlungsoption, von der sie moralisch betrachtet die besten Folgen erwartet. Eine andere Handlung wählen hieße, eine Verschlechterung des Wertintegrals gegenüber der moralisch gebotenen Handlung zu bewirken. Da der Konsequentialismus den Wert der Handlungsfolgen zum ausschließlichen Kriterium seiner Verpflichtungsurteile macht, ist man nicht nur für die Folgen seiner Handlungen verantwortlich, sondern auch für die Wertdifferenz dieser Folgen gegenüber jeder anderen potentiellen eigenen Handlung. Eine konsequentialistische Theorie kann aus diesem Grund keine Unterscheidung zwischen Handlungen und Unterlassungen vornehmen[5].

Dieses Problem der überbordenden Verantwortlichkeit beschränkt sich jedoch nicht nur auf solche Fälle, in denen eine andere Person die negativen Folgen dann bewirkt, wenn eine bestimmte – prima facie unmoralische Handlung – nicht ausgeführt wird. Vielmehr kann der Einwand Williams' wesentlich verstärkt werden: *Jede rigide unpersönliche ethische Theorie ist mit der Integrität der Person unvereinbar,* das gilt z. B. auch für eine rigide deontologische Theorie[6]. Nicht jede Theorie konsequentialisti-

[3] Sicherlich hängt das mit der methodologischen Einstellung B. Williams' zusammen, vgl. *Moralischer Zufall,* Königstein/Ts. 1984, S. 9: „Die Moralphilosophie bedarf gewiß des Nutzens von Theorie in anderen Teilen der Philosophie. Ich bin mehr denn je der Überzeugung, daß das, was sie nicht braucht, eine selbständige Theorie ist. Weder kann es eine sehr interessante, reinliche oder in sich geschlossene Theorie darüber geben, was Moralität ist, noch kann es, trotz der kraftvollen Betätigung einiger gegenwärtiger Praktiker, eine Theorie im Sinne einer philosophischen Struktur geben, die, zusammen mit einem gewissen Maß an empirischen Fakten, ein Entscheidungsverfahren für die moralische Überlegung liefert."

[4] Williams (1979) S. 72.

[5] Zur Frage des moralischen Unterschiedes zwischen Tun und Unterlassen vgl. R. Spaemann, Nebenfolgen als moralisches Problem, *Philosophisches Jahrbuch* 82 (1975) 323–335.

[6] Eine ethische Theorie ist rigide, wenn sie für jede Entscheidungssituation eindeutige Handlungsanweisungen gibt. K. Rogerson, „Williams and Kant on Integrity", *Dialogue* 22

scher Ethik hat, wie wir gesehen haben, eine interpersonell invariante Wertfunktion (und gehört damit zu den rigiden unpersönlichen ethischen Theorien). *Der strikte ethische Konsequentialismus ist allerdings mit der Integrität der Person unvereinbar.*

Betrachten wir eine beliebige strikt-konsequentialistische Theorie T mit der Wertfunktion W_T. Es ist nicht erforderlich, eine nähere inhaltliche Bestimmung von W_T vorzunehmen. Es genügt die Annahme, daß die Maximierung von W_T als *primäre* und *universale* Handlungsmotivation einer Person ungeeignet ist. Diese Annahme liegt für die utilitaristische Variante strikt-konsequentialistischer Theorien auf der Hand, aber sie ist auch für beliebige andere inhaltliche Charakterisierungen von W_T einsichtig. W_T konstituiert sich aufgrund primärer Ziele. Eine ideale moralische Person, deren ausschließliche Handlungsmotivation die Maximierung von W_T ist, kann darüber hinaus keine primären Ziele mehr verfolgen, wenn ihr nur hinreichend oft Handlungsoptionen offenstehen, die mit ihren primären Zielen nicht in Einklang stehen, aber W_T maximieren. Zur Begründung der These, daß der strikte Konsequentialismus mit der Integrität der Person unvereinbar sei, reicht daher die – für heutige menschliche Gesellschaften sicherlich zutreffende – empirische Annahme aus, daß ein Verhalten, welches eine interpersonell invariante Wertfunktion maximiert, mit der Verfolgung längerfristiger primärer Ziele nicht vereinbar ist. Selbst wenn abstrakte Ziele, wie persönliches Glück oder Zufriedenheit, als primäre Handlungsmotivation (primäres Ziel) geeignet wären, bleibt das Argument bestehen: Es ist ausgeschlossen, daß eine Person ihr persönliches Glück verwirklicht, wenn sie gezwungen ist, jede Handlungsoption zu ergreifen, die die gesellschaftliche Summe des Glücks maximiert.

Da abstrakten Zielen wie Glück als Handlungsmotivation in intentione recta eine Tendenz zur Selbstaufhebung innewohnt[7], verschärft sich das Problem rigider unpersönlicher Ethiken erheblich. Weniger abstrakte, aber als primäre Handlungsmotivation geeignetere Ziele haben i. allg. einen Zeithorizont, der eine punktuelle Verwirklichung ausschließt. Solche Ziele können eigenorientiert sein (z. B. beruflicher Erfolg), die Interessen anderer im Auge haben (z. B. die der eigenen Kinder oder Freunde) oder auf das öffentliche Wohl gerichtet sein (wie z. B. das Engagement in einer Bürgerinitiative). Sie können kurzfristig oder langfristig sein und in unterschiedlicher Stärke die Persönlichkeit des Handelnden prägen, in jedem Fall ist es wenig wahrscheinlich, daß solche Ziele eine Chance haben, verwirklicht zu werden, wenn jede Handlungswahl unter das Kriterium der unpersönlichen Folgenmaximierung fällt.

Fortsetzung Fußnote von Seite 90

(1983) 461–478 argumentiert m. E. zutreffend, daß die Ausweitung dieser Kritik auf den Typus der Kantischen Ethik unangemessen sei.

[7] Vgl. Kap. 13.

Die überbordende Theorie der Verantwortlichkeit des strikten Konsequentialismus zerstört die Integrität einer Person zunächst in dem allgemeinen Sinne, daß sie der Ausprägung einer persönlichen Handlungsorientierung zu wenig Raum läßt. Strikt-konsequentialistisch motiviertes Verhalten reduziert die Person auf ein Instrument (unpersönlicher) Wertmaximierung und ist damit unter fast[8] allen empirischen Bedingungen mit der Integrität der Persönlichkeit unvereinbar[9].

Gleiches gilt natürlich auch für rigide deontologische Theorien. Eine deontologische Theorie, deren Handlungsregel – z. B. in Gestalt eines Universalisierbarkeitskriteriums – in jeder Situation genau eine Handlung als moralisch geboten auszeichnet, kann von einer Person i. allg. ebenfalls nur unter Preisgabe ihrer persönlichen Ziele erfüllt werden. Dieser Typus einer deontologischen Theorie ist zwar theoretisch denkbar, aber entspräche nicht den gängigen Intentionen deontologischer Moraphilosophie. Für deontologische Theorien ist die Unterscheidung eines moralisch relevanten und eines moralisch irrelevanten Entscheidungsbereiches ebenso typisch wie die ethische Eingrenzung zulässiger Handlungen auf einen Bereich, innerhalb dessen nach Neigung und Klugheit entschieden werden darf.

§ 27 *Sittlichkeit*

Während es diese allgemeine Form der Unvereinbarkeit von striktem Konsequentialismus und Integrität gibt, läßt sich darüber hinaus eine sehr spezifische Form der Unvereinbarkeit aufzeigen, die man pointiert als „Zerstörung der Sittlichkeit" bezeichnen könnte.

Neben persönlichen Zielen spielen für die Ausformung einer Persönlichkeit sittliche Einstellungen eine Rolle: Bestimmte Handlungsweisen sind mit einem schlechten Gewissen verbunden, andere erwecken Abscheu oder Empörung. Die eigenen sittlichen Einstellungen prägen den Umgang mit anderen Personen, das gesellschaftliche Engagement, persönliche Zielsetzungen und die Lebensgestaltung insgesamt. Wie alle primären Ziele kommen auch diejenigen, welche auf sittlichen Einstellungen beruhen, in einen Konflikt mit der Maximierung einer unpersönlichen Wert-

[8] Eine denkbare Ausnahme wäre die extrem „moralistische" Person, deren primäres Ziel die Erfüllung bestimmter Normen ihrer subjektiven Moral darstellt. Ein mit der Bereitschaft der Selbstaufopferung einhergehender Moralismus könnte u. U. persönliche Ziele anderer Art so sehr zurücktreten lassen, daß die vollständige Erfüllung eines strikt-konsequentialistischen Handlungskriteriums ohne Zerstörung der Persönlichkeit denkbar wäre. Aber auch der ethische Konsequentialismus kann sich nicht wünschen, daß sich die Gesellschaft aus Moralisten dieser Art zusammensetzt – schon aus dem einfachen theoretischen Grunde, daß dann die kollektive Wertfunktion mangels primärer Ziele nicht mehr bestimmt werden könnte.

[9] S. Scheffler, *The Rejection of Consequentialism*, Oxford 1982, Kap. 2 versucht diesem Einwand durch eine abgeschwächte Variante des Konsequentialismus zu entgehen, die dem einzelnen das Recht einräumt, unter bestimmten Bedingungen die unpersönliche Wertfunktion der konsequentialistischen Theorie nicht zu maximieren (vgl. II.3).

funktion. In einer konsequentialistischen Gesellschaft sind sittliche Einstellungen und die darauf gründenden Ziele und Projekte durch das potentielle Handeln anderer Personen in besonderer Weise gefährdet.

Die kriteriale Berücksichtigung der de-facto-Wirkungen des je individuellen Handelns unterminiert unter plausiblen empirischen Bedingungen jede sittliche Einstellung auch dann, wenn diese Wirkungen durch das Handeln anderer Personen vermittelt sind. Diese These gilt, genau besehen, nicht nur für strikt-konsequentialistische Ethiken, sondern für jede konsequentialistische Ethik und sogar für die konsequentialistische Rationalitätskonzeption generell. Die empirischen Bedingungen für diese These sind in jeder mobilen, von Kokurrenz geprägten Gesellschaft erfüllt. Denn unter gewöhnlichen gesellschaftlichen Umständen sind die meisten Aktivitäten in der einen oder anderen Weise ersetzbar. Was der eine nicht zu unternehmen bereit ist, tut ein anderer. Zur Illustration der These seien vier Beispiel genannt:

1) Monika ist dagegen, daß in der Innenstadt alle Gehwege zugeparkt werden, also sucht sie jeweils lange nach einem Parkplatz. Monika weiß jedoch, daß die wenigen Möglichkeiten, auf den Gehsteigen zu parken, fast immer genutzt werden, so daß ihr Verzicht nur dazu führt, daß jemand anderes seinen Wagen auf den Gehsteig stellt.

2) W. kann als hoher deutscher Diplomat verhindern, daß bestimmte Pläne der Nationalsozialisten zur Durchführung kommen. Sein Amt kann W. jedoch nur behalten, wenn er bereit ist, einige der politischen Verbrechen mitzutragen und die übrigen schweigend zu dulden.

3) Georg ist Pazifist. Er hat Chemie studiert und sucht dringend einen Job. Man bietet ihm eine Anstellung in einem Forschungslabor für chemische Kriegsführung an. Georg weiß, daß dieses Labor weiterbestehen und diese Stelle ein anderer einnehmen wird, wenn er sie nicht annimmt.

4) Pedro, Armeeangehöriger in einer südamerikanischen Kleinstadt, will zehn Indianer, die sich nichts haben zuschulden kommen lassen, erschießen, um in dieser Umgebung vor weiteren Protesten gegen die Regierung abzuschrecken. Jim kommt zufällig des Weges und wird von Pedro aufgefordert, einen der Indianer zu erschießen, die übrigen würden dann freigelassen[10].

In allen diesen Beispielen ist es konsequentialistisch erlaubt und in den letzten drei Beispielen sogar geboten, entgegen der eigenen sittlichen Einstellung zu handeln. Das wäre nicht weiter beunruhigend, wenn es sich dabei um bestimmte Einzelfälle handelte, in denen die Forderungen der ethischen Theorie nicht vollständig im Einklang mit den sittlichen Normen

[10] Die letzten beiden Beispiele sind die verkürzte Fassung zweier Beispiele von B. Williams, *Kritik des Utilitarismus*, Frankfurt/M. 1979, Kap. 3. Sie wurden hier deshalb angeführt, weil sie in der ethischen Diskussion der letzten Jahre eine gewisse Prominenz erlangt haben.

sind. Die Beispiele zeigen jedoch weit mehr, nämlich *die in einer konsequentialistischen Gesellschaft epidemische Ausbreitung eines Verhaltenstypus, der sich auf das „Argument des kleineren Übels" stützt.* Dieser Verhaltenstypus ist nicht an ein strikt-konsequentialistisches Entscheidungskriterium gebunden, sondern für konsequentialistische Akteure jeder Art charakteristisch.

10. Kapitel: Individuelle Rechte

Die Probleme der Integrität, der individuellen Rechte und der Koordination sind systematisch eng miteinander verwoben. Insbesondere ist die Sicherung individueller Rechte ein wesentlicher Aspekt der Wahrung persönlicher Integrität. Die hier vertretene These lautet, daß eine konsequentialistische Ethik weder mit einer hinreichenden Wahrung persönlicher Integrität noch mit der angemessenen Sicherung individueller Rechte vereinbar ist.

§ 28 *Primäre und sekundäre Regeln*

Primäre Regeln verbieten und gebieten, verleihen den Mitgliedern einer Rechtsgemeinschaft Rechte und erlegen ihnen Pflichten auf. *Sekundäre Regeln* setzen fest, durch welches Verfahren primäre Regeln legitimerweise zustandekommen[1]. Die Grenzen zwischen beiden sind fließend. So faßt Hart das Vertragsrecht unter sekundären Regeln, weil es festsetzt, auf welche Weise neue rechtliche Verpflichtungen entstehen. Da (um bei diesem Beispiel zu bleiben) das Zustandekommen der Regeln, die den korrekten Vertragsschluß oder das Einhalten von Verträgen leiten, wiederum auf sekundäre Regeln rekurriert, ist es zweckmäßig, diese Unterscheidung nicht absolut, sondern relativ zu verstehen: Es gibt nicht zwei disjunkte Klassen von Regeln: primäre und sekundäre – sondern eine bestimmte Regel (oder ein bestimmtes Regelsystem) ist sekundär bezüglich einer anderen Regel (bzw. einem Regelsystem). So ist das Vertragsrecht sekundär gegenüber der rechtlichen Verpflichtung, die ein konkretes Vertragsverhältnis mit sich bringt, bestimmte Rechtsregeln des Grundgesetzes sind sekundär gegenüber dem Regelsystem des Vertragsrechtes etc.

Die klassische Version des Rechtspositivismus kennt keine Verpflichtung, die über eine Nötigung durch Sanktionen hinausgeht. Hart wendet dagegen ein, daß eine Rechtsregel sich von einem bloßen Befehl dadurch unterscheidet, daß sie *normativ* sei, also eine Forderung an die Adressaten richtet, die über die angekündigte Sanktion hinaus Gültigkeit hat. Ihre Gültigkeit kann darauf beruhen, daß sie *inhaltlich*, d. h. als Verhaltensregel für sich genommen, akzeptiert ist (Gültigkeit$_1$), und sie kann darauf beruhen, daß sie *aufgrund eines akzeptierten Systems sekundärer Regeln* zustandegekommen ist (Gültigkeit$_2$)[2]. Gesellschaften, die ausschließlich primäre Regeln kennen, haben nach Hart kein Rechtssystem. Für Harts Begriff des

[1] Diese Unterscheidung hat der Rechtsphilosoph H. L. A. Hart eingeführt, um den Unterschied zwischen sanktionenbewehrten Regeln und gültigem Recht gegen den traditionellen Rechtspositivismus geltend zu machen, *The Concept of Law*, Oxford 1961; dt. *Der Begriff des Rechts*, Frankfurt/M. 1973. Zur Kritik des Positivismus und seiner moderner Varianten vgl. O. Höffe, *Politische Gerechtigkeit, Grundlegung einer kritischen Philosophie von Recht und Staat*, Frankfurt/M. 1987, bes. S. 110–192.

[2] In Harts Formulierung: Eine Regel ist bindend, weil sie akzeptiert ist oder weil sie gültig ist.

Rechts sind sekundäre Regeln konstitutiv[3]. Diese rechtsphilosophisch interessante These ist für uns insofern relevant, als sich mit Hilfe dieser Unterscheidung eine zentrale Schwäche konsequentialistischer normativer Theorien deutlich machen läßt.

Betrachten wir zur Illustration folgende Situation: Eine Person P hat einer anderen Person Q zum Zeitpunkt t ein Versprechen gegeben, das P verpflichtet, eine bestimmte (generische) Handlung H zu vollziehen. P ist unterdessen (zum Zeitpunkt t') zu der festen Überzeugung gelangt, es sei für Q besser, wenn er (P) auf die Ausführung von H verzichtet – andere Interessen seien nicht tangiert: ¬H habe nach Einschätzung von P die besseren Konsequenzen für Q. Die moralische Intuition besagt, daß P nicht in allen Fällen, in denen er dieser Überzeugung ist, berechtigterweise auf die Ausführung von H verzichten darf oder gar verzichten sollte. Und darüber hinaus: Es gibt Situationen, in denen diese Bedingungen erfüllt sind und die Entscheidung von P für ¬H ein individuelles Recht von Q verletzt. Diese Intuition ist – zumindest prima facie – mit konsequentialistischer Ethik unvereinbar.

Genauer betrachtet müssen eine Reihe von Fallunterscheidungen vorgenommen werden. Eine Unterscheidung ist die zwischen einer bloß subjektiven Moral und objektiven Pflichten[4]. So könnte es immerhin sein, daß man P zwar subjektiv – angesichts seiner falschen Annahmen – für „entschuldigt" halten würde, wenn er H nicht ausführt, aber zugleich die Meinung vertritt, es sei objektiv falsch gewesen, H nicht auszuführen. Diese Unterscheidung von objektiver und subjektiver Moralität soll jedoch hier keine Rolle spielen. Daher wollen wir annehmen, daß ¬H tatsächlich die besseren Konsequenzen für Q hat, die handelnde Person dies weiß und andere Interessen nicht tangiert sind. Aber auch unter dieser Annahme scheint es unter den genannten Bedingungen nicht immer (moralisch) zulässig zu sein, wenn P H nicht ausführt. Die Begründung, dies verhalte sich deshalb so, weil P ein Versprechen gegeben habe und man Versprechen halten solle, außer wenn gravierende Gründe dagegen sprechen, ist im Rahmen des Konsequentialismus natürlich unzulässig; denn das Einhalten des Versprechens ist (wie im ersten Kapitel dargestellt wurde) für sich genommen keine Konsequenz, die für die Maximierungsforderung des Konsequentialismus eine Rolle spielen darf. Eine erfolgversprechendere konsequentialistische Argumentationsstrategie verweist auf die kau-

[3] Hart spricht i.allg. von nur *einer* sekundären rechtskonstitutiven Regel und nennt diese „rule of recognition". Eine „rule of recognition" charakterisiert das Verfahren, das eine primäre Regel in unserem Sinne zu einer gültigen$_2$ macht. Nur aufgrund dieser radikalen Vereinfachung kann Hart an der absoluten Interpretation der Unterscheidung von primären und sekundären Regeln festhalten. Eine sekundäre Regel kann in Harts Sinne nie gültig sein, da sie eine „letzte" Regel darstellt, die nur insoweit bindet, als sie akzeptiert ist. Der Rechtspositivismus Harts schlägt sich darin nieder, daß die „rule of recognition" die einzige Regel in einem Rechtssystem ist, die nur insofern bindet, als sie akzeptiert ist.

[4] „Subjektiv" hier im starken Sinne und „objektiv" im schwachen Sinne, vgl. § 23.

salen Wirkungen dieses Bruches eines Versprechens. So könnte das Vertrauen zwischen P und Q so sehr erschüttert sein, daß zukünftige Abmachungen nur noch eingeschränkt möglich sind. Wenn die nachteiligen Folgen des Vertrauensverlustes jedoch mit einbezogen werden, könnte die Ausführung von H erneut konsequentialistisch geboten sein.

Die moralische Intuition ist jedoch hartnäckig „deontologisch". Sie besteht darauf, daß die Einhaltung sittlicher Normen auch in den Fällen, in denen dies konsequentialistisch rational ist, dadurch allein nicht begründet sei. Die Einhaltung eines gegebenen Versprechens kann gute Konsequenzen haben, aber das sollte für eine moralische Person nicht der Grund sein, es einzuhalten. Um diese moralische Intuition zu testen, muß man sich nur eine Situation vor Augen führen, in denen negative Folgen der genannten Art ausgeschlossen sind. Nehmen wir zu diesem Zwecke zusätzlich an, daß P und Q sich nie mehr begegnen werden. Das Vertrauen, das Q in P gesetzt hatte, würde dann zwar ebenfalls enttäuscht, aber es wirkte sich mangels Gelegenheit nicht in Form von Koordinationsproblemen aus. Die moralische Intuition bleibt eindeutig: Auch für diesen Fall ist es P nicht immer erlaubt, H zu unterlassen, wenn P zu t' zutreffend meint, ¬ H habe die besseren Konsequenzen für Q, und P zu t Q versprochen hat, H zu tun.

Der Konsequentialist könnte seine Argumentationsstrategie fortsetzen, indem er auf die negativen Folgen dieser Enttäuschung für Qs zukünftige Lebensführung verweist. Die Enttäuschung könnte Q veranlassen, auch anderen Personen gegenüber mißtrauisch zu werden und damit seine Bereitschaft zur Koordination untergraben, was für ihn persönlich und für andere nachteilige Folgen hätte. Wie aber würde man die Situation beurteilen, wenn auch diese negativen Folgen ausgeschlossen wären? Die moralische Intuition bleibt dabei: Auch dann wäre es für P unter den genannten Bedingungen nicht immer moralisch zulässig, H zu unterlassen. Tatsächlich ist eine systematische Rekonstruktion dieser deontologischen Intuition mit Hilfe der Unterscheidung primärer und sekundärer Regeln möglich.

Institutionelle und konventionelle Regelungen entstehen im allgemeinen in einem langen historischen Prozeß, in dem rationale Entscheidungen nur eine untergeordnete Rolle spielen. Die Tradition der normativen Vertragstheorie hat dennoch hartnäckig an der historischen Fiktion der rationalen Wahl institutioneller Regelungen festgehalten. In jüngerer Zeit allerdings dominiert die Kantische Fassung des Vertragsargumentes, für die der Vertrag ein bloßes Kriterium des Rechten, also weder historische Realität noch Verpflichtungsgrund ist[5]. Die Attraktivität des Vertragsge-

[5] Besonders explizit ist dies bei Rawls (1971), ähnliches gilt jedoch auch für andere moderne Vertragstheoretiker, wie D. Gauthier, *Morals by Agreement*, Oxford 1986, J. Buchanan, *The Calculus of Consent*, Ann Arbor 1974 und R. Nozick, *Anarchy, State and Utopia*, Oxford 1974. Für den kriterialen Gebrauch des Vertragsgedankens vgl. Verf. „Der Vertragsgedanke

dankens für die politische Philosophie ergibt sich aus der besonderen Rolle, die bestehende institutionelle und konventionelle Regelungen für die moralische Handlungsorientierung spielen. Weitgehend unabhängig davon, auf welche Weise konkrete institutionelle und konventionelle Regelungen entstanden sind, d.h. unabhängig davon, ob sie den Charakter menschlicher Erfindungen haben oder in einem langen inter- und intrakulturellen Selektionsprozeß überlebt haben, ob sie demokratisch legitimiert und kontrolliert oder autoritär diktiert sind, werden sie – außer in Zeiten revolutionärer Veränderungen oder bürgerkriegsähnlicher Situationen – als *sekundäre Regeln* wahrgenommen: Sie sind Geltungsgrund für primäre Regeln und Entscheidungen erster Ordnung. Zu den primären Regeln in dem oben angeführten Beispiel gehört die „Verleihung" eines individuellen Rechtes auf H, das Q als *Resultat* des Versprechens von P zukommt; zu den primären Regeln gehört auch die *Pflicht* für P, H auszuführen. Die Institution des Versprechens selbst hat dagegen den Status einer *sekundären* Regel: Man ist verpflichtet, Versprechen einzuhalten.

Diese Unterscheidung wird deutlicher bei institutionellen Regelungen, die durch eine *Entscheidung* den Charakter eines Geltungsgrunds erhalten. Betrachten wir folgendes Beispiel: Eine Gruppe von Personen möchte gemeinsam eine Bergwanderung unternehmen. Als erfahrene Bergwanderer wissen sie, daß immer mit überraschenden Wetterumschwüngen und in Folge mit gefährlichen Situationen gerechnet werden muß. Für diesen Fall vereinbaren sie, daß einer von ihnen – wir nennen ihn Karl – die notwendigen Entscheidungen, die für die gesamte Gruppe gefällt werden müssen, übernimmt. Als die gefährliche Situation dann auch tatsächlich eintritt, entscheidet Karl, daß sich die Gruppe über einen Steilhang abseilen solle, um schnell ins Tal zu kommen. Peter, ein anderes Gruppenmitglied, ist fest davon überzeugt, daß diese Entscheidung falsch ist, da ihre Risiken weit höher sind als ein gemächlicher und unter Umständen sich bis weit in die Nacht hinziehender Abstieg über die Nordseite des Berges. Da die Gruppe solidarisch genug ist, niemanden im Stich zu lassen, könnte Peter allein durch seine strikte Weigerung erreichen, daß die Gruppe seinem Vorschlag folgt. Offensichtlich besteht ein gravierender Unterschied zwischen einer Situation, in der sich die Gruppenmitglieder zuvor auf ein bestimmtes Verfahren kollektiver Entscheidungsfindung geeinigt haben, und einer Situation, in der das nicht der Fall ist. In einer von demokratischen Werten stark geprägten Kultur würde eine Gruppe möglicherweise die Mehrheitswahlregel anwenden, wenn sie sich zuvor nicht auf ein anderes Entscheidungsverfahren geeinigt hätte, aber auch das wäre als eine sekun-

Fortsetzung Fußnote von Seite 97

in der politischen Philosophie", *Zeitschrift für Politik* 34 (1987) 200–206, sowie ders., „Plädoyer für eine kontraktualistische Philosophie der Politik" in: *Philosophie des Rechts, der Politik und der Gesellschaft*, hg. P. Koller/A. Schramm/O. Weinberger, Wien 1988, S. 51–58.

däre Regel anzusehen. Besteht eine solche sekundäre Regel nicht, wäre Peter moralisch berechtigt, ja vielleicht sogar verpflichtet, durch seine Weigerung und unter Apell an das Mitgefühl der anderen, die in seinen Augen beste kollektive Entscheidung herbeizuführen. Gibt es jedoch eine (allgemein akzeptierte) sekundäre Regel, z. B. diejenige, daß Karl die kollektiven Entscheidungen – soweit notwendig – zu treffen hat, dann müßte Peter ein Verhalten dieser Art erst noch begründen – prima facie wäre es nicht akzeptabel. Primäre Regeln können sekundäre brechen: Wenn Peter tatsächlich fest davon überzeugt ist, daß ein Abstieg, wie ihn Karl bevorzugt, zum sicheren Tod der Gruppe führt, ist er möglicherweise sogar verpflichtet, die aufgrund einer allgemein akzeptierten Regel kollektiver Entscheidungsfindung zustandegekommene kollektive Entscheidung zu torpedieren.

Legitimität ist ein Merkmal kollektiver Entscheidungen, die im Einklang mit bestimmten sekundären Regeln zustandegekommen sind. Legitimität heißt nicht, daß diese Entscheidungen inhaltlich (im Sinne eines Systems primärer Regeln) richtig sind. Nicht jede legitime Entscheidung ist deshalb (unter einem moralischen Aspekt) richtig, aber es gilt auch: Nicht jede moralisch falsche Entscheidung ist illegitim. Übertragen auf den Adressaten kollektiver Entscheidungen: Auch eine falsche, aber legitime kollektive Entscheidung kann verpflichten, aber nicht jede falsche, aber legitime kollektive Entscheidung verpflichtet. Der Rechtspositivismus normativer Art vertritt hingegen die These: Eine legitime kollektive Entscheidung verpflichtet immer, entweder weil es keine falschen legitimen Entscheidungen gibt, da ein unabhängiger Maßstab des Rechten außerhalb eines Systems sekundärer Regeln nicht existiert, oder weil der unabhängige Maßstab im Sinne eines „relativen" Naturrechts keinen Verpflichtungscharakter hat. Der (normative) Rechtspositivismus bildet das eine Extrem auf einer Skala, an deren anderem Extrem die Auffassung angesiedelt ist, sekundäre Regeln seien moralisch irrelevant. Konsequentialismus und Existenzialismus treffen sich in der gemeinsamen Ablehnung sekundärer Regeln – der Existenzialismus explizit, der Konsequentialismus nur implizit als logische Folge seiner spezifischen Rationalitätskonzeption.

§ 29 *Individuelle Rechte als Beschränkungen*

Konventionen, Institutionen und sittliche Bindungen sind durch je spezifische sekundäre Regeln charakterisiert. Personen, die gemeinsam einer bestimmten gesellschaftlichen Ordnung angehören, verhalten sich in hohem Maße im Einklang mit sekundären Regeln der in dieser gesellschaftlichen Ordnung etablierten Konventionen, Institutionen und sittlichen Bindungen. Das gilt auch für dasjenige Verhalten, das im Sinne des ersten Kapitels Handlungscharakter hat. Darüber hinaus scheint die Befolgung

sekundärer Regeln nicht immer, aber in vielen Fällen moralisch geboten zu sein. Die Frage, die sich uns stellt, lautet daher: Kann ein Verhalten im Einklang mit sekundären Regeln als konsequentialistisch rational gelten? Selbstverständlich kann es in einer gegebenen Situation eine Verhaltensmöglichkeit geben, die im Einklang mit den je anzuwendenden sekundären Regeln und zugleich konsequentialistisch rational ist – insofern könnte diese Frage trivial mit „ja" beantwortet werden. Gemeint ist jedoch folgendes: Gibt es in jeder Entscheidungssituation, in der sekundäre Regeln $R_1, R_2, \ldots, R_n$ eine Rolle spielen und die Einhaltung dieser Regeln moralisch geboten ist, eine Handlungsoption, die $R_1, R_2, \ldots, R_n$ erfüllt, und die zugleich konsequentialistisch geboten ist? Zwei Fälle sind dabei zu unterscheiden: konsequentialistisch geboten bezüglich der subjektiven Wertfunktion des Handelnden (eine Handlung, die diese Bedingung erfüllt, ist konsequentialistisch rational) und konsequentialistisch geboten bezüglich der für die moralische Beurteilung wesentlichen Wertfunktion (eine Handlung, die diese Bedingung erfüllt, ist im Sinne des ethischen Konsequentialismus moralisch geboten).

Es ist für eine konsequentialistische Ethik nicht erforderlich nachzuweisen, daß jede Verhaltensweise, die im Einklang mit bestimmten, für die Situation relevanten sekundären Regeln ist, auch konsequentialistisch rational ist, da es natürlich auch unmoralisches und irrationales Verhalten gibt, das mit bestimmten sekundären Regeln in Einklang ist. *Der Konsequentialismus in der Ethik wäre jedoch sicherlich dann inadäquat, wenn sich aufzeigen ließe, daß es Situationen gibt, in denen es moralisch geboten ist, im Einklang mit bestimmten sekundären Regeln zu handeln, diese Handlungen aber für alle Wertfunktionen, die als Grundlage der moralischen Beurteilung in Betracht kommen*[6]*, konsequentialistisch irrational sind.* Es scheint auf den ersten Blick völlig ausgeschlossen zu sein, diesen komplexen Sachverhalt, der eine Existenzbehauptung mit einem Allsatz verbindet, zu beweisen oder zu widerlegen. Tatsächlich läßt sich jedoch die These

Es gibt Situationen, in denen einige der – aufgrund sekundärer Regeln – moralisch gebotenen Handlungen für keine zulässige Wertfunktion$_m$ konsequentialistisch rational sind

unter plausiblen normativen Prämissen präzise beweisen.

Da sich die zu beweisende These nicht auf alle, sondern nur auf einige Situationen bezieht, können wir eine nähere Charakterisierung vornehmen: Wir betrachten ausschließlich Situationen, in denen es keine externen Effekte der offenstehenden Handlungen (außerhalb einer Personengruppe K) gibt, und wir gehen vereinfachend davon aus, daß jede vollständige Kombination individueller Handlungen und jede kollektive

[6] Wertfunktionen, die als Grundlage der *moralischen* Beurteilung in Betracht kommen, bilden eine Teilmenge aller für die *subjektive* Beurteilung möglichen Wertfunktionen. Wertfunktionen dieser Art kennzeichnen wir durch „$_m$".

Handlung[7] einen bestimmten Nachfolgezustand festlegten[8], der Element der Alternativenmenge X ist. Wenn externe Effekte ausgeschlossen sind, ist es naheliegend, für jede auf Situationen dieser Art anwendbare Wertfunktion$_m$ *Paretoinklusivität* zu fordern.

DEF.: Eine Aggregationsregel f ist paretoinklusiv (P) genau dann, wenn gilt

$$\forall g \in G: \forall x,y \in X: [\forall i \in K: \langle x,y\rangle \in g(i) \wedge$$
$$\exists j \in K: \langle x,y\rangle \in \dot{g}(j)] \rightarrow \langle x,y\rangle \in \dot{f}(g)]^9$$

Die Forderung der Paretoinklusivität ist eine äußerst schwache Bedingung, unabhängig von der Interpretation, die man den verwendeten Grundbegriffen gibt. In der üblichen subjektivistischen Interpretation gibt g die Struktur der subjektiven Präferenzen der Personen aus K wieder. Eine Regel kollektiver Entscheidungsfindung ist jedenfalls dann paretoinklusiv, wenn ein gesellschaftlicher Zustand aus X, der von allen Personen aus K vorgezogen wird, auch in der kollektiven Bewertung besser ist (die o.g. Definition ist etwas schwächer). Der Ausdruck „kollektive Bewertung" darf dabei nicht mißverstanden werden: In einer vollkommenen marktwirtschaftlichen Konkurrenzökonomie werden ausschließlich pareto-optimale Güterverteilungen erreicht, damit ist die Regel kollektiver Entscheidungsfindung, wie sie durch eine vollkommene, marktwirtschaftliche Konkurrenzökonomie realisiert ist, paretoinklusiv. Die kollektive Bewertung, wie sie in R zum Ausdruck kommt, ist daher nicht notwendigerweise mit einem personalen Träger dieser Bewertung verknüpft – es kann durchaus sein, daß es in K niemanden gibt, der die durch R repräsentierte kollektive Bewertung teilt. Auch jede subjektivistische ethische Theorie läßt sich durch eine Aggregationsfunktion darstellen, auch wenn dazu bisweilen zusätzliche Begriffe, insbesondere eine erweiterte Informationsbasis, erforderlich ist. So kann man das utilitarische Kriterium der Durchschnittsnutzen-Maximierung nur mit Hilfe vollständiger interpersoneller Einheiten-Vergleichbarkeit der individuellen Bewertungen wieder-

[7] Ein gegebener gesellschaftlicher Zustand geht in einen anderen gesellschaftlichen Zustand durch Veränderungen über, die i.allg. nicht ausschließlich durch menschliche Handlungen hervorgerufen sind. Dieser Einfluß könnte auch in diesem Modell jederzeit durch eine probabilistische Komponente berücksichtigt werden. Nur aus Gründen der Einfachheit wird hier darauf verzichtet. Eine Handlung wurde in § 4 als Übergang zu einem Nachfolgezustand im Weltenbaum interpretiert. Wenn Personen handeln, ergibt sich eine dadurch eindeutig bestimmte Abfolge von Weltzuständen.

[8] Um dieses Bild für den n-Personenfall zu vereinfachen, stellen wir uns eine Abfolge von Weltzuständen (hier, da auf K bezogen, von *gesellschaftlichen* Zuständen) vor, deren Verzweigungen nicht durch eine individuelle Handlung, sondern durch ein n-Tupel individueller Handlungen, bzw. von einer kollektiven Handlung bestimmt werden. Die Unterlassung einer Handlung gilt dabei ebenfalls als Handlung, so daß auch der individuellen Handlung – ceteris paribus – diejenige kollektive Handlung entspricht, bei der ausschließlich die betreffende Person eine Veränderung des gesellschaftlichen Zustandes bewirkt.

[9] Die hier verwendete Terminologie wurde in § 20 eingeführt.

geben. Auch nicht-subjektivistische Interpretationen der Pareto-Inklusivität sind möglich. Zum Beispiel kann man die individuellen Relationen g(i) als Repräsentation objektiver individueller Vorteile ansehen: Genau dann, wenn eine Alternative x für eine Person i objektiv vorteilhafter ist als eine Alternative y, gilt $\langle x,y\rangle \in g(i)$, und zwar unabhängig von der subjektiven Überzeugung der betreffenden Person i. Pareto-Inklusivität hieße hier, daß ein Zustand x, der für alle vorteilhafter ist als ein Zustand y, auch realisiert wird, wenn diese beiden Alternativen offenstehen.

Nicht nur die neuzeitliche liberale Tradition des politischen Denkens, sondern – wenn auch in weniger pointierter Form – fast jeder Ansatz der praktischen Philosophie berücksichtigt, daß Personen – aus naturrechtlichen, anthropologischen oder aus anderen Gründen – ein gewisses Maß an persönlicher Freiheit, ein persönlicher Freiheitsspielraum, zusteht. Individuelle Rechte sind Ausdruck dieser Freiheit. Es ist moralisch betrachtet trivial, daß es Entscheidungssituationen gibt, in der einzelne Personen das Recht haben, für sich zu entscheiden, welche Alternative sie verwirklichen wollen. Dieses Recht ist nicht unbedingt, es gilt nicht für beliebige kontingente Annahmen, ja vermutlich ist jedes Recht unter bestimmten kontrafaktischen Annahmen obsolet. Jeder hat zum Beispiel das Recht zu entscheiden, ob er auf dem Rücken oder auf dem Bauch schlafen möchte, um ein viel zitiertes Beispiel von A. K. Sen[10] anzuführen, aber wenn meine Entscheidung, auf dem Bauch zu schlafen, eine Explosion in einer nahegelegenen Stadt auslösen würde, wäre dieses prima-facie-Recht aufgehoben. Es ist an dieser Stelle jedoch in keiner Weise notwendig, sich Gedanken über die Bedingungen individueller Rechte zu machen. Es reicht hin, eine Minimalanforderung zu formulieren, die von jeder Wahrnehmung individueller Rechte in Anspruch genommen würde. Eine Person, die ein individuelles Recht hat, muß zumindest über eine Alternative frei entscheiden können, bzw. eine Person, die ein individuelles Recht wahrnimmt, entscheidet zumindest über eine Alternative.

Die Minimalanforderung für die Wahrnehmung individueller Rechte läßt sich als eine Eigenschaft der Aggregationsregel formulieren: Die kollektive Relation R muß die individuelle Entscheidung der betreffenden Person bezüglich mindestens einer Alternative x,y respektieren. Wir nennen diese Minimalanforderung deshalb auch „Liberalitätsbedingung" (L).

$$\forall g \in G: \forall i \in K: \underset{x \neq y}{\exists} x,y \in X\,[(\langle x,y\rangle \in \dot{g}(i) \rightarrow$$
$$\langle x,y\rangle \in \dot{f}(g)) \wedge (\langle y,x\rangle \in \dot{g}(i) \rightarrow \langle y,x\rangle \in \dot{f}(g))]$$

Es läßt sich zeigen, daß die Liberalitätsbedingung und die Forderung der Pareto-Inklusivität logisch unvereinbar sind: *Es gibt keine Aggregations-*

[10] A. K. Sen, *Collective Choice and Social Welfare*, London u.a. 1970, Kap. 6.

regel, die für beliebige Präferenzstrukturen paretoinklusiv ist und die Liberalismusbedingung erfüllt[11].

Beweis[12]: Für $\#K \geq 2 \wedge \#X \geq 2$;

Seien $x_1, x_2, x_3, x_4 \in X$; $A, B \in K$; $A \neq B$.

1. Fall: $\langle x_1, x_2 \rangle = \langle x_3, x_4 \rangle$ (1)

angenommen $\langle x_1, x_2 \rangle \in \dot{g}(A)$ und $\langle x_3, x_4 \rangle \in \dot{g}(B)$ (2)

$\langle x_1, x_2 \rangle \in \dot{f}(g)$ wg. (2) und L (3)

$\langle x_2, x_1 \rangle \in \dot{f}(g)$ wg. (1), (2) und L (4)

(Widerspruch) wg. (3) und (4) (5)

2. Fall: $\{x_1, x_2\}$ und $\{x_3, x_4\}$ haben ein gemeinsames Element. (6)

o. B. d. A. sei $x_1 = x_3$.

[11] Dieser Zusammenhang wurde von A. K. Sen 1970 entdeckt und als Liberales Paradoxon bezeichnet (vgl. *Collective Choice and Social Welfare*, London u.a. 1970, Kap. 6 u. 6*, Sen „Liberty, Unanimity and Rights, *Economica* 43 (1976) 217–145 u. ders. „Personal Utilities and Public Judgements or What's Wrong with Welfare Economics?, *Journal of Economy* 84 (1979) 537–558 sowie K. Suzumura, *Rational Choice, Collective Decisions and Social Welfare*, Cambridge 1983, Kap. 7 und J. L. Wriglesworth, *Libertarian Conflicts in Social Choice*, Cambridge 1985. Eine sehr klare, aber nicht-technische Darstellung gibt J. Barnes, „Freedom, Rationality, and Paradox" *Canadian Journal of Philosophy* 10 (1980) 545–565. Die Literatur zu diesem Theorem ist unterdessen gewaltig angeschwollen, ohne daß eine durchweg befriedigende „Lösung" im Rahmen der Logik kollektiver Entscheidungen gefunden wurde. Die Interpretation, die wir diesem Theorem geben, führt zu der Schlußfolgerung, daß es im Rahmen einer konsequentialistischen Rationalitätskonzeption keine Lösung geben kann. Das Liberale Paradoxon kann man demnach als eine (besonders elegante) Widerlegung eines konsequentialistischen Rationalitäts- und Moralverständnisses verstehen. Zur Diskussion im Rahmen der Logik kollektiver Entscheidungen vgl. die Generalisieurng des Theorems durch D. Austen-Smith, „Restricted Pareto and Rights, *Journal of Economic Theory* 26 (1982) 89–99; D. Kelsey, „The Liberal Paradox – A Generalisation, *Social Choice and Welfare*, 1 (1985) 245–250. Einen Versuch, das Problem im konsequentialistischen Rahmen zu lösen, unternimmt A. Gibbard, „A Pareto-Consistent Libertarian Claim, *Journal of Economic Theory* 7 (1974) 388–410 – modifiziert von J. Kelly, „Rights Exercising and Pareto-Consistent Libertarian Claim, *Journal of Economic theory* 13 (1976) 138–153 und widerlegt von K. Suzumura, „Liberal Paradox and the Voluntary Exchange of Rights-Exercising, *Journal of Economic Theory* 22 (1980) 407–422 – erneut C. Mezzetti, „Paretian Efficiency, Rawlsian Justice and Nozick Theory of Rights, *Social Choice and Welfare* 4 (1987) 25–37. Andere Theoretiker haben, veranlaßt durch dieses Theorem, den konzeptionellen Rahmen der Logik kollektiver Entscheidungen verlassen, ohne aber eine nicht-konsequentialistische Alternative zu entwickeln, so z. B. C. K. Rowley und A. T. Peacock, *Welfare Economics: A Liberal Restatement*, London 1975; Farell (1976) und R. Sudgen, „Social Choice and Individual Liberty", in: M. J. Artis und A. R. Noboy (Hg.), *Contemporary Economic Analysis*, London 1978. J. Aldrich, „The Dilemma of a Paretian Liberal. Some Consequences of Sen's Theorem, *Public Choice* 30 (1976) 1–20 hat gezeigt, daß es einen logischen Zusammenhang zwischen diesem Theorem und der bekannten Interaktionsstruktur des Gefangenendilemmas gibt, auf die wir erst in Teil IV eingehen. Vgl. auch L. Kern/J. Nida-Rümelin: Eine Lösung des liberalen Paradoxons, in: Verf. (Hg.), *Praktische Rationalität*, Berlin 1993.

[12] Der Beweis stimmt inhaltlich mit dem des Sen-Theorems, Sen (1970), a.a.O. überein. Da Sen jedoch eine andere (und wegen der Identifikation von Funktion und Wertebereich nicht ganz saubere) Terminologie verwendet, wird hier ein Beweis gegeben, der dem formal ungeübten Leser durch Zeilenverweise das Verständnis erleichtern soll.

angenommen: $\langle x_1,x_2\rangle \in \dot{g}(A) \wedge \langle x_4,x_1\rangle \in \dot{g}(B) \wedge$
$\forall i \in K: \langle x_2,x_4\rangle \in \dot{g}(i)$ (7)

$\langle x_1,x_2\rangle \in \dot{f}(g)$ wg. (7) und L (8)

$\langle x_4,x_1\rangle \in \dot{f}(g)$ wg. (7) und L (9)

$\langle x_2,x_4\rangle \in \dot{f}(g)$ wg. (7) und P (10)

f ist nicht azyklisch wg. (8), (9) u. (10). (11)

3. Fall: $\# \{x_1,x_2,x_3,x_4\} = 4$ (12)

angenommen: $\langle x_1,x_2\rangle \in g(A) \wedge \langle x_3,x_4\rangle \in g(B) \wedge$
$\forall i \in K [\langle x_4,x_1\rangle \in \dot{g}(i) \wedge \langle x_2,x_3\rangle \in \dot{g}(i)]$ (13)

$\langle x_1,x_2\rangle \in \dot{f}(g)$ wg. (13) und L (14)

$\langle x_3,x_4\rangle \in \dot{f}(g)$ wg. (13) und L (15)

$\langle x_4,x_1\rangle \in \dot{f}(g)$ wg. (13) und P (16)

$\langle x_2,x_3\rangle \in \dot{f}(g)$ wg. (13) und P (17)

f ist nicht azyklisch wg. (14), (17), (15) u. (16). (18)

Anmerkung: Die Annahmen (2), (7) und (13) sind zulässig, da eine kollektive Wohlfahrtsfunktion nach Definition beliebigen Präferenzstrukturen g kollektive Präferenzrelationen mit Ordnungseigenschaften zuordnet. Es genügt also, jeweils für eine Präferenzstruktur aufzuzeigen, daß die Aggregationsregel keine kollektive Wohlfahrtsfunktion sein kann, wenn die Liberalitätsbedingung und das Pareto-Kriterium erfüllt sind[13].

Die ursprüngliche Interpretation des Theorems war, daß die bloße Berücksichtigung von Nutzenaspekten bei der kollektiven Bewertung von Alternativen mit der Gewährung individueller Rechte bzw. Entscheidungsfreiheiten nicht vereinbar sei[14], daß vielmehr auch Handlungsfolgen, die nicht ausschließlich durch ihren Beitrag zur Nutzenverteilung charakterisiert sind, berücksichtigt werden müssen. Tatsächlich gibt es einen interessanten logischen Zusammenhang zwischen bloßer Präferenzorientierung, Paretobedingung und Irrelevanzbedingung[15], der für die weitere Diskussion wesentlich ist.

Um diesen Zusammenhang aufzuzeigen, benötigen wir einige Definitionen:

[13] Beispiele zur Illustration des Theorems finden sich in L. Kern, „Lösungen des Liberalen Paradoxes", in: *Analytische Politiphilosophie und ökonomische Rationalität*, hg. von K.-P. Markl, Bd. I, Opladen 1985.

[14] Vgl. A. K. Sen, „Personal Utilities and Public Judgements or What's Wrong with Welfare Economics?", *Journal of Economics* 84 (1979) 537–558.

[15] Vgl. dazu D. Kelsey: „The Liberal Paradox – Generalisation, *Social Choice and Welfare* 1 (1985) 245–250, ergänzend D. A. MacIntyre, „‚The Liberal Paradox: A Generalisation' by D. Kelsey", *Social Choice and Welfare* 4 (1987) 219–223 und J. Riley, „Generalized Social Welfare Functionals: Welfarism, Morality and Liberty", *Social Choice and Welfare* 3 (1986) 233–254.

DEF.: Eine Aggregationsregel f ist eine kollektive Entscheidungsfunktion (KEF) gdw. $\forall g \in G$ [f(g) ist azyklisch].

DEF.: Eine Aggregationsregel f ist eine kollektive Wohlfahrtsfunktion (KWF) gdw. $\forall g \in G$ [f (g) ist transitiv].

DEF.: Eine Aggregationsregel f erfüllt die Bedingung $\overline{I}$ gdw. die kollektive Präferenz über eine Alternative x,y aus X ausschließlich von den individuellen Präferenzen bez. x und y abhängt (bzw.: wenn die kollektive Präferenz eine *Funktion* der individuellen Präferenzen ist).

DEF.: Eine Aggregationsregel erfüllt die Bedingung NI (Non-Imposition) der kollektiv freien Wahl gdw.

$\forall x,y \in X: \exists g \in G [\langle x,y\rangle \in \dot{f}(g)]$.

Die Bedingung NI ist die Präzisierung einer intuitiv plausiblen Minimalbedingung eines jeden demokratischen Entscheidungsverfahrens: Die kollektive Wahl soll von den individuellen Präferenzen in irgendeiner Weise – in welcher kann hier offenbleiben – abhängen oder, mit anderen Worten: Die kollektive Wahl darf nicht völlig unabhängig davon sein, was die betroffenen Personen wünschen.

Eine leichte Abschwächung der Bedingung lautet:

DEF.: Eine Aggregationsregel f erfüllt die schwache Bedingung NI_s der kollektiv freien Wahl gdw.

$\forall x,y \in X: \exists g \in G [\langle x,y\rangle \in f(g)]$.

Daß diese Bedingung recht schwach ist, läßt sich schon daraus ersehen, daß sie durch die universale kollektive Indifferenz als kollektiver Präferenzordnung erfüllt wäre.

DEF.: Eine Aggregationsregel f nennen wir *präferenzenorientiert* gdw.

$\forall g \in G [\forall i \in K^{+}_{x,y} [\langle v,w\rangle \in \dot{g}(i)] \wedge \forall i \in K^{-}_{x,y} [\langle w,v\rangle \in \dot{g}(i)] \wedge$
$\forall i \in K^{0}_{x,y} [\langle v,w\rangle \in \tilde{g}(i)] \rightarrow (\langle v,w\rangle \in f(g) \leftrightarrow \langle x,y\rangle \in f(g))]$.

Wobei: $K^{+}_{x,y} := \{i | i \in K \wedge \langle x,y\rangle \in \dot{g}(i)\}$,
$K^{-}_{x,y} := \{i | i \in K \wedge \langle y,x\rangle \in \dot{g}(i)\}$,
$K^{0}_{x,y} := \{i | i \in K \wedge \langle x,y\rangle \in \tilde{g}(i)\}$.

Die Mengen $K^{+}_{x,y}$, $K^{-}_{x,y}$, $K^{0}_{x,y}$ bilden für beliebige Alternativenpaare x,y ∈ X wegen der Ordnungseigenschaften (Vollständigkeit, Reflexivität und Transitivität) der individuellen Präferenzrelationen eine Partition von K.

Eine präferenzorientierte Aggregationsregel berücksichtigt andere Informationen als solche über die vorliegende Struktur individueller Präferenzen nicht: Gleiche Präferenzstrukturen führen unter präferenzenorientierten Aggregationsregeln zu gleichen (kollektiven oder moralischen) besser-Relationen. Das (schwache) Pareto-Prinzip impliziert für den Fall, daß

die Präferenzen der Personen übereinstimmen, Präferenzenorientierung im Sinne dieser Definition (es impliziert nicht generell Präferenzenorientierung). Präferenzenorientierung kann man daher auch als ein *Merkmal* des Pareto-Prinzips ansehen. Dieses Merkmal der bloßen Berücksichtigung individueller Präferenzen bei der Bestimmung des moralisch Gebotenen ist ein wichtiger Aspekt der Unvereinbarkeit von Pareto-Kriterium und individuellen Rechten (Liberalitätsbedingung).

Eine genauere Betrachtung der logischen Zusammenhänge ergibt folgendes Bild: Eine Aggregationsregel, die die Rationalitätsbedingung der *Transitivität* der moralischen besser-Relation sowie die Irrelevanzbedingung erfüllt und zugleich jede moralische Rangordnung zuläßt, soweit nur entsprechende individuelle Präferenzen vorliegen (NI_s), ist präferenzenorientiert.

Theorem:

Eine KWF, die die Bedingungen NI_s und I erfüllt, ist präferenzenorientiert.

Beweis:

Angenommen: $K^{+}_{x,y} = K^{+}_{v,w} \wedge K^{-}_{x,y} = K^{-}_{v,w} \wedge K^{0}_{x,y} = K^{0}_{v,w}$ (1)

Angenommen: $\# \{x,y,v,w\} = 4$ (2)

Angenommen: $\langle x,y\rangle \in \dot{f}(g)$ (3)

Angenommen: $g \in G\,[\langle v,x\rangle \in f(g) \wedge \langle y,w\rangle \in f(g)]$ (4)

wg. NI_s gibt es ein solches g

$\langle v,w\rangle \in \dot{f}(g)$ wg. (3) u. (4) (5)

f (g) transitiv, da KWF

Angenommen: $g \in G\,\langle w,x\rangle \in f(g) \wedge \langle y,v\rangle \in f(g)$ (6)

wg. NI_s gibt es ein solches g

$\langle w,v\rangle \in \dot{f}(g)$ wg. (6) u. (3) (7)

(7) ist jedoch mit (1), (3) und I nicht vereinbar, damit ist der Fall (6) (und aus dem gleichen Grund die übrigen Fälle) ausgeschlossen, bzw. gezeigt, daß (3) → (5). (8)

Angenommen: $\langle v,w\rangle \in \dot{f}(g)$ (9)

Angenommen: $g \in G \wedge [\langle x,v\rangle \in f(g) \wedge \langle w,y\rangle \in f(g)]$ wg. NI_s (10)

$\langle x,y\rangle \in \dot{f}(g)$ wg. (10) u. (9) (11)

Angenommen: $g \in G \wedge [\langle y,v\rangle \in f(g) \wedge \langle w,x\rangle \in f(g)]$ (12)

$\langle y,x\rangle \in \dot{f}(g)$ wg. (12) u. (9) (13)

(13) ist jedoch mit (1), (9) u. I unvereinbar, damit ist der Fall (12) ausgeschlossen, bzw. gezeigt, daß (9) → (11). (14)

Die übrigen Fälle $\#\{x,y,v,w\}=3$ und $\#\{x,y,v,w\}=2$ können in analoger Weise gezeigt werden.

Während eine kollektive Wohlfahrtsregel, die die Bedingungen NI_s und I erfüllt, präferenzenorientiert ist, läßt sich beweisen, daß es keine vernünftige Aggregationsregel gibt, die die Bedingung NI_s erfüllt und zugleich individuelle Rechte berücksichtigt. Sen hat gezeigt, daß reine Präferenzenorientierung mit der Berücksichtigung individueller Rechte unvereinbar ist. Jetzt kann diese Aussage genauer qualifiziert werden: Der logische Konflikt besteht schon zwischen der Bedingung NI_s und der Liberalitätsbedingung. Es gibt keine kollektive Wohlfahrtfunktion, die zugleich die Liberalitätsbedingung und NI_s erfüllt.

Wie immer ein System individueller Rechte inhaltlich charakterisiert ist, in jedem Fall muß die Minimalbedingung L erfüllt sein. Dieser Sachverhalt gilt unabhängig davon, durch welche formellen oder informellen Verfahren kollektive Entscheidungen getroffen werden: durch Abstimmungen, Aushandeln, institutionelle Autorität oder unabhängige Einzelentscheidungen. Jedem gesellschaftlichen Prozeß kann in einem formalen (oder weiteren Sinne) eine „kollektive Entscheidung" zugeordnet werden, auch wenn dieser keine „kollektiven Entscheidungen" im Sinne eines etablierten Verfahrens kollektiver Entscheidungsfindung beinhaltet, das berechtigen würde, von einer „Entscheidung der betreffenden Personengruppe" zu sprechen. Kollektive Entscheidungen im weiteren Sinne manifestieren sich in bestimmten realisierten gesellschaftlichen Zuständen. In der gesellschaftlichen Realität spielen für die Ausprägung eines gesellschaftlichen Zustandes nicht nur individuelle und im engeren Sinne „kollektive Entscheidungen", sondern auch probabilistische Prozesse und zufällige Ereignisse eine Rolle. Diesen probabilistischen Aspekt müssen wir ausklammern, aber im übrigen kann man sich kollektive Entscheidungsprozesse i. w. S. als beliebig komplex vorstellen – sie lassen sich jeweils in Gestalt einer *Aggregationsregel* (f) darstellen, die *Strukturen individueller Präferenzen* (g) eine *kollektive Präferenzrelation* (R) zuordnet. Da es darüber hinaus nicht erforderlich ist, die individuellen Präferenzrelationen g (i) subjektivistisch als tatsächliche (sich im Entscheidungsverhalten manifestierende) „Präferenzen der betreffenden Person" zu interpretieren, erhalten die oben dargelegten logischen Zusammenhänge ein großes Gewicht. Wenn die individuellen Relationen g (i) z. B. nicht die möglicherweise entstellten oder einfach „unvernünftigen" oder „antisozialen" subjektiven Präferenzen wiedergeben, sondern als Repräsentation des individuellen Wohls interpretiert werden, dann ist die Bedingung der Pareto-Inklusivität ein notwendiger Bestandteil konsequentialistischer Ethik, auch dann, wenn man den Bereich konsequentialistischer Ethik nicht auf subjektivistische Theorien einschränkt. Soweit die Liberalitätsbedingung (L) als Minimalbedingung individueller Rechte akzeptiert wird, gibt es also Entscheidungssituationen, in denen es moralisch geboten ist, im Einklang

mit sekundären Regeln zu handeln, die sicherstellen, daß individuelle Rechte nicht verletzt werden, und dieses Handeln ist für alle Wertfunktionen, die als Grundlage der moralischen Beurteilung in Betracht kommen, konsequentialistisch irrational. Mit anderen Worten: Es gibt Situationen, in denen einige der – aufgrund sekundärer Regeln – moralisch gebotenen Handlungen für keine zulässige Wertfunktion konsequentialistisch rational sind. Die Ausgangsthese dieses Abschnittes ist unter den genannten Annahmen bestätigt.

Dieses Ergebnis ist für die Kritik des Konsequentialismus von zentraler Bedeutung. Es enthebt uns der Aufgabe, im einzelnen zu prüfen, ob bestimmte normative Bedingungen mit konsequentialistischer Ethik vereinbar sind: *Jede adäquate Berücksichtigung individueller Rechte ist in bestimmten Entscheidungssituationen mit konsequentialistischer Ethik unvereinbar.*

Der ethische Konsequentialist könnte dagegen zwei Argumente vorbringen: (a) L sei keine notwendige Bedingung der Berücksichtigung individueller Rechte, oder (b) der logische Konflikt sei *praktisch irrelevant,* da die in der konkreten Anwendung zu berücksichtigenden individuellen Relationen g (i) eine Gestalt haben, die diesen logischen Konflikt (von L und PO) de facto nicht wirksam werden läßt. (b) kann in eine normative und eine empirische Fassung aufgespalten werden: Die normative Fassung (b1) fordert aus moralischen Gründen, daß die individuellen Präferenzen bzw. Wertfunktionen eine solche Gestalt haben sollten, während die empirische Fassung (b2) diese Gestalt der individuellen Präferenzen als de facto gegeben annimmt.

In der Tat kann die Liberalitätsbedingung (L) nicht den Anspruch erheben, eine inhaltliche Charakterisierung individueller Rechte zu sein. *L ist weder eine für alle Entscheidungssituationen notwendige noch hinreichende Bedingung der angemessenen Berücksichtigung individueller Rechte.* L ist nicht in allen Entscheidungssituationen *notwendig,* weil es einen großen Bereich von Entscheidungssituationen gibt, in denen individuelle Rechte auch dann nicht verletzt sind, wenn L nicht erfüllt ist. Zu diesem Bereich gehören insbesondere öffentliche, z. B. parlamentarische Entscheidungen: Bei der Verabschiedung von Gesetzen wird keinem einzelnen Parlamentarier die Möglichkeit eingeräumt, über mindestens eine Gesetzesalternative (kollektiv, d. h. hier bindend) zu entscheiden, ohne daß deshalb von einer Verletzung individueller Rechte gesprochen werden kann. L ist aber auch nicht *hinreichend,* weil es zur Sicherung individueller Rechte in vielen Fällen nicht ausreicht, der einzelnen Person das Entscheidungsrecht bezüglich nur *einer* Alternative einzuräumen. Darüber hinaus wäre L auch dann erfüllt, wenn sich ein von der Aggregationsregel eingeräumtes (formales) individuelles Entscheidungsrecht auf Alternativen bezieht, die unter einem ethischen Aspekt nicht zu den berechtigten individuellen Ansprüchen zu zählen sind.

Obwohl L nicht in allen Entscheidungssituationen zur Sicherung individueller Rechte notwendig ist, *gilt, daß* es bezüglich jedes *adäquaten Systems individueller Rechte Entscheidungssituationen gibt, in denen L eine notwendige Bedingung der Berücksichtigung individueller Rechte darstellt.* Das erste Argument ist damit entkräftet.

Die empirische Fassung des zweiten Argumentes (b2) könnte durch eine Erhebung wohl rasch zurückgewiesen werden, die normative Fassung (b1) hat jedoch prima facie eine gewisse Plausibilität; denn damit die Berücksichtigung individueller Rechte (Bedingung L) mit der konsequentialistischen Optimierung in Konflikt gerät, muß man individuelle Präferenzen annehmen, die ein auffälliges Charakteristikum gemeinsam haben: Es handelt sich jeweils um individuelle Präferenzen, die sich auf Alternativen beziehen, deren Auswirkungen im wesentlichen eine andere Person betreffen. Präferenzen, die diese Eigenschaft haben, könnte man als *„einmischende"* oder *„paternalistische"* Präferenzen bezeichnen. Die für den logischen Konflikt notwendige partielle Übereinstimmung der individuellen Präferenzen (Bedingung P) beruht charakteristischerweise nicht auf eigenorientierter Optimierung, sondern auf der Rücknahme eigener Interessen zugunsten vermuteter und vielleicht nur vermeintlicher („objektiver") Interessen anderer Personen. Der logische Konflikt konsequentialistischer Optimierung und individueller Rechte ist ausgeschlossen, wenn die individuellen Präferenzen ausschließlich eigenorientiert sind[16].

Eigenorientierung ist nur *eine* Möglichkeit der Nichteinmischung. Es ist auch eine *moralisch motivierte* Nichteinmischung denkbar. Wenn jede der Personen zwischen allen Alternativen, die sie zu dem persönlichen Entscheidungsbereich einer anderen beteiligten Person zählt, aus Rücksichtnahme indifferent wäre und diese Alternativenmengen interpersonell übereinstimmen, kommt es ebenfalls nicht zu einem Konflikt zwischen konsequentialistischer Optimierung und der Berücksichtigung individueller Rechte. Der ethische Konsequentialist kann daher argumentieren, das Problem der angemessenen Berücksichtigung individueller Rechte sei ohne Aufgabe des konsequentialistischen Ansatzes zu lösen, indem man ausschließlich solche individuellen Präferenzen (und individuellen Wertfunktionen) berücksichtigt, die einen *nicht-paternalistischen* Charakter haben. Diese Verteidigung des Konsequentialismus fordert jedoch einen hohen Preis: Sie schränkt in drastischer Weise die Präferenzensouveränität des Einzelnen ein, wodurch zugleich die Grundlagen konsequentialistischer Ethik unterminiert werden. Es ist ein wesentliches Element konsequentialistischer Rationalität, daß die vorausgehende Absicht durch eine Handlung erfüllt wird, die *im Urteil des Handelnden* optimale Folgen hat. Die zur Umgehung des Konfliktes zwischen individuellen Rechten und konsequentialistischer Maximierung notwendigen Beschränkungen verän-

[16] Diese These wird in L. Kern/J. Nida-Rümelin, „Eine Lösung des Liberalen Paradoxons" in: *Praktische Rationalität* hg. v. Verf. bewiesen.

dern jedoch die zu maximierende Wertfunktion in einer Weise, die ausschließt, daß sie nach wie vor die subjektiven Bewertungen der handelnden Person repräsentiert: *Um den Konflikt zu umgehen, muß verlangt werden, daß der Einzelne auf die Realisierung eines Zustandes verzichtet, der von ihm bei Abwägung aller Merkmale der Entscheidungssituation für besser gehalten wird. Diese Einschränkung ist jedoch nicht nur mit einem teleologischen, sondern auch mit einem deontologischen Konsequentialismus unvereinbar. Individuelle Rechte lassen sich im konsequentialistischen Rahmen nicht angemessen berücksichtigen.*

11. Kapitel: Das Koordinationsproblem

Das Problem der konsequentialistischen Rekonstruktion sekundärer Regeln und speziell der Berücksichtigung individueller Rechte ist Teil des Koordinationsproblems des Konsequentialismus generell. Insofern kann man die Diskussion des letzten Abschnittes als eine Illustration der folgenden, zunächst abstrakten Darstellung des Koordinationsproblems ansehen.

§ 30 Die ideale moralische Person

Wir haben schon mehrfach von der idealen moralischen Person bezüglich einer Theorie T gesprochen. Eine solche Person erfüllt alle sittlichen Anforderungen, die T stellt – unabhängig davon, was Gegenstand dieser Anforderungen ist (konkrete Handlungen bei „handlungsorientierten" Theorien, Dispositionen bei Tugendethiken, Maximen etc.). Die ideale moralische Person bezüglich T bezeichnen wir als P_T. In der Geschichte der Moralphilosophie wurde eine Vielzahl von Gegenstandsarten einem moralisch wertenden Urteil unterzogen. Dazu gehören neben den oben genannten Handlungen, Maximen und Dispositionen auch Handlungsregeln, Institutionen und Güterverteilungen. Für alle denkbaren Gegenstände der moralischen Beurteilung gibt es allerdings ein Primat der Handlung; denn für jede ethische Theorie ist die *normative* (die verpflichtende) Komponente zentral. Verpflichtungsurteile können sich jedoch nur auf solche Gegenstände beziehen, für die eine moralische Verantwortlichkeit besteht („Aus Sollen folgt Können"). Verantwortlichkeit besteht jedoch streng genommen nur für Handlungen und davon abgeleitet für ihre Eigenschaften (z. B. Regelkonformität) und ihre vorhersehbaren Folgen. Werturteile, die nicht den Charakter von Verpflichtungsurteilen haben, erhalten ihre *ethische Relevanz* durch ihre Beziehung zu Verpflichtungsurteilen. Ein System von Werturteilen, das keinerlei Implikationen für die Bestimmung des Pflichtgemäßen hat (z. B. weil der Gegenstand der wertenden Beurteilung von menschlichem Handeln nicht beeinflußt werden kann), ist kein sinnvoller Bestandteil einer ethischen Theorie. Es ist daher gerechtfertigt, wenn wir uns zunächst auf den Fall der Erfüllung einer ethischen Theorie durch Handlungen einer Person konzentrieren.

Konsequentialistische Ethiken enthalten – explizit oder implizit – eine bestimmte Wertfunktion (W). Die ideale moralische Person bezüglich einer konsequentialistischen Theorie T maximiert mit ihren Handlungen die Wertfunktion W_T im Sinne des Bayes'schen Konsequentialismus (s. §§ 8–10). Die Bayes'sche Subjektrelativierung ist zur Charakterisierung der idealen moralischen Person unverzichtbar, auch wenn die Klasse bez. T gebotener Handlungen objektiv im starken Sinne (nach § 23) bestimmt sein sollte. Welche Rolle die Bayes'sche [vgl. o.] Subjektrelativierung

spielt, wird deutlich, wenn man der idealen moralischen Person P_T eine Person P_T' gegenüberstellt, die zwar de facto die Wertfunktion W_T maximiert, aber ohne daß die entsprechenden inneren Gründe (aus § 9) die Handlungen von P_T' motivieren. P_T maximiert die Wertfunktion *intentional*, während P_T' die Wertfunktion nur *formal* maximiert. Es liegt auf der Hand, daß diese Unterscheidung sich im wesentlichen mit der Kantischen zwischen der *Legalität* (P_T') und der *Moralität* (P_T) einer Handlung deckt.

§ 31 *Bedingungen der Kooperativität einer Theorie*

Eine konsequentialistische Theorie T wollen wir „*im schwachen Sinne kooperativ*" nennen (sie erfüllt die *Bedingung der schwachen Kooperativität*), wenn gilt, daß eine abgeschlossene Gruppe von idealen moralischen Personen bez. T die Wertfunktion W_T maximiert. Eine „abgeschlossene" Gruppe besteht dabei aus Personen, die in keiner kausalen oder moralischen[1] Beziehung zu Personen außerhalb der Gruppe stehen.

Wenn T eine kooperative Theorie (im schwachen Sinne) ist, dann maximiert das Handeln einer abgeschlossenen Gruppe von P_T die Wertfunktion W_T. Wenn alle Menschen ideale moralische Personen bez. einer (schwach) kooperativen Theorie sind, dann hat ihr Handeln die (im Sinne der Theorie) bestmöglichen Konsequenzen, d. h. es gibt in diesem Fall keine andere Kombination (oder „Struktur") individuellen Handelns, das bessere Konsequenzen hätte. Punktuell formuliert: Wenn alle Personen in jeder Entscheidungssituation T erfüllen und T die Bedingung K erfüllt, dann gibt es zu keinem Zeitpunkt eine andere mögliche Kombination individuellen Handelns, das bessere Konsequenzen hätte[2].

Statt nur zu fordern, daß *in einer idealen moralischen Welt* die Übereinstimmung jedes Einzelnen mit den Kriterien der Theorie auch im Sinne der Theorie selbst wünschenswert ist, kann man auch als „*starke Kooperativitätsbedingung*" fordern, daß eine (beliebige) Gruppe von idealen moralischen Personen *unter allen Umständen* W_T maximiert. Insbesondere maximiert eine Gruppe von P_T auch dann die Wertfunktion einer *stark kooperativen* Theorie, wenn andere an der Interaktion Beteiligte keine idealen moralischen Personen (bez. T) sind. Im Spezialfall der Einergruppe impliziert die Bedingung der starken Kooperativität, daß jede Person, wenn sie sich an die normativen Kriterien einer stark kooperativen Theorie hält, die Wertfunktion dieser Theorie („global", also bei Berücksichtigung aller für die Bewertung wesentlichen Merkmale der Situation) maximiert. Eine stark kooperative Theorie ist daher konsequentialistisch, aber nicht jede konsequentialistische Theorie ist (stark) kooperativ. Selbst-

[1] Eine moralische, aber vermutlich nicht kausale Beziehung besteht z. B. zwischen einem Verstorbenen und seinen Angehörigen. Die Angehörigen werden bei der moralischen Bewertung den Willen des Verstorbenen auch dann berücksichtigen, wenn sie annehmen können, daß ihre Entscheidungen keine kausalen Auswirkungen für den Verstorbenen haben.

[2] Vgl. D. Regan, *Utilitarianism and Co-operation*, Oxford 1980, Kap. 1 u. Kap 6.

verständlich impliziert starke Kooperativität einer Theorie schwache Kooperativität.

Eine strikt-konsequentialistische ethische Theorie zeichnet sich durch eine interpersonell invariante Wertfunktion aus. Im Gegensatz zu anderen, schwächeren Formen konsequentialistischer Theorien treten für strikt-konsequentialistische ethische Theorien Koordinationsprobleme nicht auf, wie sie paradigmatisch durch den Interaktionstyp des sog. Gefangenendilemmas (vgl. § 45) beschrieben werden. Handlungsutilitaristen sind daher in Versuchung, das Koordinationsproblem als ein bloßes Problem der Anwendung anzusehen, das auf theoretischer Ebene durch die Invarianzeigenschaft der Wertfunktion gelöst ist und sich nur im Rahmen *nicht-idealer Anwendungsbedingungen* stellt[3]. Es läßt sich jedoch zeigen, daß diese Betrachtungsweise nicht aufrechtzuerhalten ist. Der wesentliche Grund dafür ist, daß eine interpersonell invariante Wertfunktion das Koordinationsproblem zwar in gewisser Hinsicht unter einem statischen, aber keinesfalls unter einem dynamischen Aspekt obsolet macht. Aber auch in der statischen Betrachtung bleibt selbst im Rahmen des strikten Konsequentialismus ein Koordinationsproblem, dem wir uns zunächst zuwenden müssen.

§ 32 *Das statische Koordinationsproblem des strikten Konsequentialismus*

Die These dieses Abschnittes lautet: *Auch bei invarianten Bewertungsfunktionen können Interaktionssituationen auftreten, in denen konsequentialistisch moralisches Handeln die Wertfunktion nicht maximiert.*

Angenommen, eine abgeschlossene Gruppe besteht aus zwei Personen P und P'. Wir wollen weiter annehmen, daß beiden Personen jeweils drei Handlungsalternativen offenstehen (für P: a,b,c, und für P': a',b',c'). Eine Interaktionssituation dieser Art hat einen Gleichgewichtspunkt (z. B. in a und b'), wenn weder für P noch für P' eine der anderen Alternativen (b und c, bzw. a' und c') besser gewesen wäre, sofern man die Handlungswahl der anderen Person als gegeben annimmt. Natürlich können auch mehrere Handlungskombinationen Gleichgewichtspunkte der Interaktionssituation darstellen[4]. *Reine Koordinationssituationen* sind dadurch ge-

[3] Vgl. etwa D. Birnbacher, *Verantwortung für zukünftige Generationen*, Stuttgart 1988, I.1.2 und Kap. 6 sowie R. M. Hare, *Moralisches Denken*, Frankfurt 1992, Kap. I.

[4] In der Spieltheorie werden Gleichgewichtspunkte auch Lösungen oder Sattelpunkte genannt. Bisweilen wird zwischen einer Lösung im strengen Sinn und einer Lösung im Sinne eines Gleichgewichtspunktes unterschieden. Eine Lösung im strengen Sinne hat zwei Merkmale: diejenige des Gleichgewichtes und zusätzlich diejenige der Pareto-Optimalität der Handlungskombination, d. h. es darf keine andere Handlungskombination geben, die alle Beteiligten – gemessen an ihren individuellen Bewertungsfunktionen – besser stellt. Im Falle einer Menge mehrerer pareto-optimaler Gleichgewichtspunkte ist jedes Element dieser Menge nur dann eine Lösung im starken Sinne, wenn die pareto-optimalen Gleichge-

prägt, daß es keinen Interessenkonflikt zwischen den Beteiligten gibt, formal schlägt sich das in interpersonell gleichen Bewertungen der Gleichgewichtskombinationen nieder. *Wenn die Beteiligten ideale moralische Personen bez. einer strikt konsequentialistischen Theorie sind, dann ist jede Interaktion eine reine Koordinationssituation.* Interaktionssituationen können konfliktfrei sein, auch ohne daß es sich bei den Beteiligten um ideale moralische Personen im Sinne des strikten Konsequentialismus handelt. Ebenso können Interaktionssituationen konfliktfrei sein, obwohl die Bewertungsfunktionen der Beteiligten interpersonell variieren. Wenn allerdings alle Interaktionen bez. einer abgeschlossenen Gruppe von Personen reine Koordinationssituationen[5] sind, dann sind die Bewertungsfunktionen interpersonell invariant, was allerdings noch nicht hinreicht, um die Beteiligten als ideale moralische Personen auszuweisen. Dazu wäre eine strikt konsequentialistische Theorie notwendig, deren Wertfunktion mit dieser (dann bis auf lineare Transformation eindeutig bestimmten) interpersonell invarianten Bewertungsfunktion identisch ist.

Bei interpersonell invarianten Bewertungsfunktionen gilt trivialerweise, daß *jeder Gleichgewichtspunkt einer Interaktionssituation zugleich für beliebige, bezüglich der individuellen Bewertungsfunktionen monotone Wertfunktionen*[6] *individuell maximierend* ist – allerdings nur „individuell maximierend"; denn es ist ja, wie gesagt, nicht ausgeschlossen, daß es auch bei interpersonell invarianten Bewertungsfunktionen mehrere Gleichgewichtspunkte gibt, die von W_T unterschiedlich bewertet werden. Unter der Bedingung, daß die anderen Personen diejenigen Handlungen wählen, die Bestandteil des jeweiligen Gleichgewichtspunktes sind, ist jede Einzelentscheidung des Gleichgewichtspunktes unter der Bedingung interpersonell invarianter Bewertungsfunktionen konsequentialistisch rational (bez. jeder monotonen Wertfunktion und damit u. a. auch bez. jeder – pareto-inklusiven – Utilitarismusvariante). Wenn es jedoch einen anderen Gleichgewichtspunkt gibt, der von W_T höher bewertet wird, dann handelt zwar jeder Beteiligte im Sinne des strikten ethischen Konsequentialismus rational, aber dennoch maximiert die Kombination dieser individuellen Handlungen W_T nicht: *Der strikte ethische Konsequentialismus erfüllt nicht die*

Fortsetzung Fußnote von Seite 113

wichtspunkte äquivalent und austauschbar sind. Zwei Gleichgewichtspunkte sind äquivalent, wenn alle Beteiligten zwischen ihnen indifferent sind, und sie sind austauschbar, wenn die individuellen Handlungen der Kombinationen, neu kombiniert, wieder einen Gleichgewichtspunkt ergeben.

[5] Eine sorgfältige Klassifikation von Koordinationsproblemen mit dem Instrumentarium der Spieltheorie gibt D. Lewis, *Konventionen. Eine sprachphilosophische Abhandlung*, Berlin/New York 1975, Kap. 2. Lewis stützt sich auf T. C. Schelling, *The Strategy of Conflict*, Cambridge/Mass. 1960. Schelling hat durch seine Analyse in den 60er Jahren zu einer Erweiterung des Anwendungsgebietes der Spieltheorie wesentlich beigetragen.

[6] Sei die Wertfunktion W der Theorie eine Funktion der individuellen Bewertungsfunktionen f_1, f_2, f_3. W ist monoton genau dann, wenn gilt, daß jede Erhöhung einer beliebigen individuellen Bewertung bei unveränderten Bewertungen der übrigen eine Erhöhung von W nach sich zieht.

Bedingung der (schwachen) Kooperativität, d. h. es gibt keine strikt konsequentialistische ethische Theorie T, für die gilt, daß jede abgeschlossene Gruppe von idealen moralischen Personen bez. T die Wertfunktion WT maximiert. Anders formuliert: Auch in einer unter moralischen Aspekten idealen Welt würden sich nicht notwendigerweise die (im Sinne der Theorie) bestmöglichen Konsequenzen einstellen. Da starke Kooperativität einer Theorie schwache Kooperativität impliziert, gilt a fortiori, daß es *keine strikt konsequentialistische ethische Theorie T gibt, für die gilt, daß jede Gruppe von idealen moralischen Personen bez. T unter allen Umständen die betreffende Wertfunktion maximiert.*

Dieses Ergebnis ist jedoch für den (strikten) Konsequentialismus nicht wirklich fatal: Erstens, weil es sich um ein *epistemisches* Problem handelt, und zweitens, weil sich überraschenderweise zeigen läßt, daß *es keine Interaktionssituation gibt, die ausschließt, daß individuell rationales Handeln im Sinne des strikten Konsequentialismus zugleich (kollektiv) maximierend ist.*

§ 33 *Die (statische) Vereinbarkeit von Kooperativität und striktem ethischen Konsequentialismus*

Die These dieses Abschnittes lautet: *Wenn sich die Bewertungsfunktionen als Repräsentation motivierender Absichten für die moralisch idealen Personen bez. einer Theorie T nicht unterscheiden, dann sind Interaktionssituationen ausgeschlossen, in denen konsequentialistisch rationales Handeln und die Maximierung der Wertfunktion unvereinbar sind.* Um Mißverständnisse auszuschließen, sei auf folgende Aspekte der These hingewiesen: a) Die Wertfunktion der Theorie hat denselben Gegenstandsbereich („Definitionsbereich") und dieselbe Gestalt wie die Bewertungsfunktionen der einzelnen Personen. Für den Fall interpersoneller Varianz wäre es erforderlich, zusätzlich eine zweistellige Wertfunktion der Theorie einzuführen, die den einzelnen Personen jeweils *ihre* gebotene Bewertungsfunktion zuweist: Interpersonelle Varianz impliziert nicht notwendigerweise eine relativistische Theorie – das unter dem Stichwort „abgeschwächter Konsequentialismus" diskutierte Gerechtigkeitskriterium G_1 ist dafür ein Beispiel. b) Interpersonelle Invarianz der Bewertungsfunktionen der P_T's impliziert nicht notwendigerweise die eine oder andere Version der Nutzenmaximierung (sei es Durchschnitts- oder Gesamtnutzenmaximierung), vielmehr ist sie mit jeder *g-konsequentialistischen* Theorie vereinbar, wobei g-konsequentialistische Theorien, obwohl sie i. allg. als *deontologisch* zu interpretieren sind, dennoch dem Typ des *strikten Konsequentialismus* angehören[7].

[7] Dies folgt unmittelbar aus den Überlegungen von §§ 18, 19.

Der Beweis der These ist einfach. Aus der Annahme des Gegenteils läßt sich ein Widerspruch ableiten. Angenommen, es gibt bei interpersonell invarianten Bewertungsfunktionen als Repräsentation motivierender Absichten eine Interaktionssituation, in der konsequentialistisch rationales Handeln und die Maximierung der Wertfunktion unvereinbar sind, obwohl die Beteiligten ideale moralische Personen sind, dann müßte es eine Kombination individueller Handlungen geben, die die Wertfunktion der Theorie maximiert, obwohl mindestens eine Person nicht konsequentialistisch rational handelt. Betrachten wir eine solche Person: Wenn sie nicht konsequentialistisch rational handelt, dann steht ihr eine Handlungsalternative offen, die bei unverändertem Verhalten der übrigen Beteiligten zu einem Zustand führt, der von der Wertfunktion der Theorie höher bewertet wird. Dies wurde jedoch von der Annahme ausgeschlossen, so daß die These bewiesen ist.

Nun sind in der Literatur zahlreiche Beispiele diskutiert worden, die die Vorzüge einer regelutilitaristischen Theorie gegenüber dem Handlungsutilitarismus aufzeigen sollen und die mit dieser These scheinbar in direktem Widerspruch stehen[8] Wenn es nicht gelingt, deutlich zu machen, daß diese Beispiele nur scheinbar zeigen, was sie zeigen sollen, dann wäre die Argumentation dieses Abschnittes nicht aufrechtzuerhalten. Dies ist ein Grund dafür, im folgenden einige dieser Standardbeispiele zu analysieren. Der zweite Grund ist, daß sich anhand dieser Beispiele wichtige Differenzierungen der bisherigen These ergeben, deren ursprüngliche Fassung lautete: Wenn sich die Bewertungsfunktionen als Repräsentation motivieren-

[8] Die Auseinandersetzung mit den unterschiedlichen Typen von Verallgemeinerungsargumenten in utilitaristischem Kontext würde hier zu weit führen. Wir beschränken uns auf einen Hinweis zur Literatur: Eine nicht-utilitaristische, auch im üblichen Sinne deontologische Theorie entwickelt W. D. Ross, *The Right and the Good*, Oxford 1930. Das „Standardwerk", welches für eine eigene Theorie normativer Ethik Pate steht, die in der angelsächsischen Diskussion vielfach als „utilitarian generalization" (im Gegensatz zu „rule utilitarianism") charakterisiert wird – im Deutschen werden beide Varianten meist als „Regelutilitarismus" bezeichnet – ist das Buch von M. G. Singer, *Generalization in Ethics. An Essay in the Logic of Ethics, with the Rudiments of a System of Moral Philosophy* (1961), dt. *Verallgemeinerung in der Ethik. Zur Logik moralischen Argumentierens*, Frankfurt/M. 1975. Das zentrale Argument des Buches ist die Ableitung des sog. Prinzips der Verallgemeinerung (eine Handlung sollte dann nicht ausgeführt werden, wenn ihre Ausführung durch alle schlechte Folgen hätte) aus dem handlungsutilitaristischen Prinzip (man sollte eine Handlung dann nicht ausführen, wenn sie überwiegend schlechte Folgen hat) und dem Prinzip der Gleichheit (Personen in gleichen Umständen unterliegen den gleichen ethischen Geboten). N. Hoerster hat sich mit diesem Argument kritisch auseinandergesetzt und bietet zugleich eine gute Einführung in die Thematik: *Utilitaristische Ethik und Verallgemeinerung*, Freiburg/München 1971, vgl. auch D. Lyons, *Forms and Limits of Utilitarianism*, Oxford 1965 und J. Narveson, *Morality and Utility*, Baltimore 1967; S. Toulmin, *The Place of Reason in Ethics*, Cambridge 1970 sowie D. Regan, *Utilitarianism and Cooperation*, Oxford 1980, Kap. 6, weitere einschlägige Diskussionsbeiträge sind: J. Harrison, „Utilitarianism, Universalisation, and our Duty to be Just", *Proc. Arist. Soc., Suppl.* (1952/53); A. C. Ewing, „What would Happen if Everbody Acted Like Me?" *Philosophy* 28 (1953) 16–29; A. K. Stout, „But Suppose Everyone Did the Same", *Autralasian J. Phil.* 51 (1954) 1–29; R. B. Brandt, „Towards a Credible Form of Rule-Utilitarianism", in: *Morality and the Language of Conduct*, hg. von G. Nakhnikian /H.-N. Castaneda, Detroit 1963.

der Absichten für die moralisch idealen Personen bez. einer Theorie T nicht unterscheiden, dann sind Interaktionssituationen ausgeschlossen, in denen (individuell) konsequentialistisch rationales Handeln und die (kollektive) Maximierung der Wertfunktion unvereinbar sind.

§ 34 *Entkräftung der Gegenbeispiele*

Ein erstes Beispiel ist die Entscheidungssituation, in der sich ein Bürger angesichts einer politischen Wahl befindet. Da eine nationale Wahl wohl noch nie durch nur eine Stimme entschieden wurde, gilt für jede einzelne Stimmabgabe, daß sie das Ergebnis (etwa die Zusammensetzung der zukünftigen Regierung oder die zukünftige Gesetzgebung) nicht beeinflußt. Dennoch legen manche Bürger Wert darauf, sich an der Wahl zu beteiligen[9]. In vielen Fällen läßt sich diese Bereitschaft problemlos strikt-konsequentialistisch rekonstruieren. So könnte es sein, daß der betreffende Bürger damit rechnen muß, von seinen Bekannten gefragt zu werden, ob er auch gewählt hat, und es mag ihm dann unangenehm sein, eingestehen zu müssen, daß er der Wahl ferngeblieben ist. Zur Wahl zu gehen könnte daher für diesen Bürger sogar im Sinne des *eigenorientierten Konsequentialismus* rational sein. Weniger plausibel ist es anzunehmen, daß die Beteiligung an der Wahl zwar nicht rational im Sinne des eigenorientierten Konsequentialismus gewesen ist, aber rational im Sinne des *Gerechtigkeitskonsequentialismus,* aber auch diese Möglichkeit ist nicht ausgeschlossen. Allerdings ist nicht garantiert, daß ein Bürger, der sich aus Gründen der Gerechtigkeit für die Wahl entschieden hat, damit auch rational im Sinne des Gerechtigkeitskonsequentialismus gehandelt hat. Der betreffende Bürger mag es als in hohem Maße ungerecht empfinden, wenn sich einige wenige ihrer Staatsbürgerpflichten entziehen, während andere ihnen auch dann nachkommen, wenn sie an diesem strahlenden Wochenende lieber Urlaub gemacht hätten, ohne daß sich ein solches Gerechtigkeitsempfinden im Rahmen des Gerechtigkeitskonsequentialismus rekonstruieren ließe. In diesem Fall spielen zwar Gerechtigkeitsaspekte eine *motivierende* Rolle, aber damit ist noch nicht garantiert, daß eine Handlung, die diese Aspekte berücksichtigt, ihre *Folgen* (unter Gerechtigkeitsaspekten) optimiert – ja im allgemeinen wird die Wahlhandlung die Gerechtigkeit des gesellschaftlichen Zustandes *nicht optimieren;* denn die singuläre Entscheidung eines Bürgers zu wählen tangiert in den meisten Fällen die Gerechtigkeit des Gesamtzustandes nicht. Eine Verteilung [10, 10, 10, 0, 0] ist nicht notwendigerweise ungerechter als eine Verteilung [10, 10, 10, 10, 0]. Für die meisten Kriterien der Verteilungsgerechtigkeit spielt die Anzahl

[9] Dieses Beispiel, auch „Wähler-Paradox" genannt, ist in der Literatur häufig diskutiert worden. Es läßt sich wohl nicht mehr feststellen, wer es als erster angeführt hat, vgl. u. a. D. Lyons, *Forms and Limits of Utilitarianism,* Oxford 1965, S. 90; B. Brody, „The Equivalence of Act and Rule Utilitarianism", *Phil. Studies* 18 (1967) 81–86, S. 82.

der Personen, die eine bestimmte Zuteilung erhalten, keine Rolle. Aus dem Sachverhalt, daß es ungerecht ist, nicht zu wählen, kann also nicht geschlossen werden, daß ein gesellschaftlicher Zustand mit der Anzahl der Nicht-Wähler – ceteris paribus – (in monoton steigender Funktion) ungerechter wird. *Die Ungerechtigkeit einer Handlung überträgt sich i. allg. nicht auf die Ungerechtigkeit ihrer Folgen, und umgekehrt ist die Ungerechtigkeit einer Handlung nicht abhängig von der relativen Ungerechtigkeit der Weltzustände, die sie kausal beeinflußt.*

Die Entscheidungssituation des Wählers wird für unsere These erst dann interessant, wenn seine Entscheidung nicht dadurch motiviert ist, daß er für den Fall der Nicht-Wahl persönlich (oder moralisch) negative Folgen gewärtigen müßte. Nach der Wahl stellt er fest, daß seine Stimme für den Wahlausgang völlig irrelevant war, da der Unterschied zwischen den Oppositions- und den Regierungsparteien über eine Million Stimmen betrug. Obwohl es in diesem Fall keine positiven Wirkungen der Handlung (Wahlbeteiligung) gibt, sprechen gute Gründe für diese Entscheidung. Der Wähler könnte seine Entscheidung z. B. dadurch rechtfertigen, daß er auf die negativen Folgen aufmerksam macht, die es hätte, wenn jeder, der weder ein intrinsisches Bedürfnis hat, sich an der Wahl zu beteiligen, noch sich Illusionen über die Wahrscheinlichkeit macht, daß gerade seine Stimme das Ergebnis beeinflussen könnte, von der Wahl absehen würde – vermutlich würde das eine Krise der demokratischen Ordnung nach sich ziehen. Halten wir fest: Die Wahlbeteiligung läßt sich unter einigen plausiblen empirischen Annahmen, wie es scheint, nicht mit dem Wert ihrer Folgen rechtfertigen, und dennoch scheint die Wahlbeteiligung unter einigen ebenso plausiblen normativen Annahmen, wenn nicht moralisch geboten, so doch zumindest moralisch erlaubt und sicherlich nicht unvernünftig zu sein; Wahlenthaltung scheint dagegen im Sinne des strikten Konsequentialismus geboten und Wahlbeteiligung verboten zu sein. Da der klassische Handlungs- ebenso wie der zeitgenössische Gerechtigkeitsutilitarismus Spielarten des strikten Konsequentialismus sind, kann dieses Ergebnis gegen beide Formen des Utilitarismus ins Feld geführt werden. In der Tat haben Beispiel dieser Art bei der Entstehung des Regelutilitarismus Pate gestanden. Mit der naheliegenden zusätzlichen Annahme, daß allgemeine Wahlbeteiligung gegenüber allgemeiner Wahlenthaltung einen höheren Rang in der intersubjektiv invarianten Wertfunktion der betrachteten strikt-konsequentialistischen Theorie einnimmt, scheint die individuelle Maximierung dieser Wertfunktion unter den genannten empirischen Voraussetzungen zu einer kollektiven Entscheidung zu führen, die diese Wertfunktion nicht maximiert. *Dies aber steht im Widerspruch zu der Ausgangsthese, daß es im Rahmen einer strikt-konsequentialistischen Theorie keine Interaktionssituation geben kann, in der (individuell) konsequentialistisch rationales Handeln und die (kollektive) Maximierung der Wertfunktion unvereinbar sind.*

Bevor wir diesen überraschenden Widerspruch auflösen, sei noch ein zweites Beispiel vorgestellt[10]: „Professor A hat bis spät in die Nacht in seinem Büro gearbeitet. Auf seinem Heimweg durch das Universitätsgelände liegt eine große, sehr gepflegte Rasenfläche, die er gewöhnlich, dem Wege folgend, umgeht. Heute jedoch kommt A der Gedanke, die vor ihm liegende Rasenfläche zu überqueren und so die Dauer seines Weges um einige Minuten abzukürzen. Das Gras würde sich schnell von seinen Schritten erholen und sicher keinen Schaden nehmen. Anders wäre es allerdings, wenn jeder Benutzer des Universitätsgeländes aus Bequemlichkeit diese Abkürzung wählte. Dann würde das Gras an der betreffenden Stelle verkümmern (...)". Ist A verpflichtet, auch dieses eine Mal auf die Überquerung des Rasens zu verzichten?

Ein drittes Beispiel: In München herrscht seit Tagen das bekannte herbstliche Inversionswetter. Kalte, aufliegende Luftschichten verhindern den Abzug der Emissionen von Industrie, Verkehr und Hausbrand. Da die Meßwerte noch nicht das gesetzliche Minium für ein Fahrverbot erreicht haben, beschränkt sich die Stadtverwaltung auf die Aufforderung, nicht das eigene Auto, sondern die öffentlichen Verkehrsmittel zu benutzen. Herr Mayer steht vor der Frage, ob er mit dem eigenen PKW zur Arbeit fahren soll.

Diese Liste der Beispiele ließe sich beliebig verlängern. Es handelt sich um alltägliche Situationen, und sie haben offensichtlich eine gemeinsame Struktur: Naheliegende Wertannahmen lassen ein Verhalten als konsequentialistisch rational erscheinen, das verallgemeinert – d. h. von beliebigen Personen unter gleichen Umständen praktiziert – schlechte Konsequenzen hat (und zwar schlecht im Sinne der gleichen Wertfunktion, die die einzelne Handlung als konsequentialistisch rational auszeichnet). Dies steht im Widerspruch mit den abstrakten Überlegungen, die dafür sprechen, daß eine solche Unvereinbarkeit individueller und kollektiver Maximierung im Rahmen *strikt-konsequentialistischer* Theorien nicht möglich ist. Tatsächlich sind diese und ähnliche Entscheidungssituationen überraschenderweise nachweisbar nur vermeintlich Beispiele für die Unvereinbarkeit individueller und kollektiver Maximierung im Rahmen *strikt-konsequentialistischer* Theorien. Eine genauere Analyse zeigt, daß die Standardinterpretation dieser Beispiele auf einer Vermengung unterschiedlicher Fragestellungen und Gegenstandsbereiche beruht.

Bevor wir darauf eingehen, muß betont werden, daß wir in keiner Weise davon Gebrauch machen wollen, daß die infrage stehenden individuellen Handlungen dem gleichen Handlungstyp angehören. Der Widerspruch zur Ausgangsthese, daß es im Rahmen einer strikt-konsequentialistischen Theorie keine Interaktionssituation geben kann, in der (individuell) konsequentialistisch rationales Handeln und die (kollektive) Maximie-

[10] Dieses Beispiel stammt von N. Hoerster, *Utilitaristische Ethik und Verallgemeinerung*. Freiburg/München ²1977, S. 43 f.

rung der Wertfunktion unvereinbar sind, ergibt sich ja vielmehr erst dann, wenn der Bereich individueller Handlungen *nicht eingeschränkt* ist. Wenn die Beispiele als Illustration des Verallgemeinerungsprinzips verstanden werden (man sollte nur solche Handlungen vollziehen, die allgemein vollzogen keine schlechten Folgen haben), dann entsteht dieser Widerspruch nicht. Die Feststellung: „Es gibt Situationen, in denen individuell maximierendes Verhalten kollektiv nicht maximierend (im Rahmen des strikten Konsequentialismus) ist, *wenn alle das gleiche tun*, steht nicht im Widerspruch mit der These: „Es gibt keine Situation, in der individuelle und kollektive Maximierung (im Rahmen des strikten Konsequentialismus) unvereinbar sind". Der logische Widerspruch verschwindet wegen der Einschränkung der individuellen Entscheidungsfreiheit („wenn alle das gleiche tun").

Damit ist das Problem jedoch keineswegs gelöst, denn die Beispiele kommen auch ohne jede Einschränkung des individuellen Entscheidungsbereiches aus. Im Wählerbeispiel hat jeder nur zwei Handlungsalternativen. Unter naheliegenden empirischen Annahmen könnte es für alle oder zumindest für einen großen Teil der Bürgerschaft strikt konsequentialistisch rational sein, sich der Wahl zu enthalten. Diese geringe Wahlbeteiligung als Ergebnis individueller Maximierung aber könnte zugleich (kollektiv) nicht maximierend sein. Das Beispiel scheint also doch eine Widerlegung der Ausgangsthese zu sein.

Betrachten wir das Beispiel etwas genauer. Die Wahlakte geschehen nicht gleichzeitig. Man stelle sich vor, es wähle einer nach dem anderen und das Stimmenverhältnis wird jeweils auf einer öffentlichen Anzeigetafel registriert. Zunächst steht es Null zu Null. In dieser Situation ist es für jeden Bürger, der eine bestimmte politische Präferenz hat, konsequentialistisch rational, seine Stimme abzugeben. Wenn einer seine Stimme für A abgegeben hat, wird ein Anhänger von B (nur A und B stehen zur Wahl) seine Stimme für B abgeben. Unter konsequentialistisch rationalen Bürgern wird sich dieses Wechselspiel so oft wiederholen, bis das Wählerpotential einer der Parteien erschöpft ist, genauer: bis es keinen Bürger mit einer bestimmten Parteipräferenz mehr gibt, für dessen Wertfunktion die Entscheidung über die Regierung mehr ins Gewicht fällt als die Vorteile seiner Wahlenthaltung. Nachdem die stärkere Gruppierung eine Stimme Mehrheit hält, werden keine weiteren Stimmen mehr abgegeben. Im Rahmen des strikten Konsequentialismus müssen wir unsere Betrachtung auf eine der beiden Gruppen (Anhänger von A bzw. Anhänger von B) beschränken, da die Parteipräferenz Bestandteil der gemeinsamen Wertfunktion ist. Jede Wahlhandlung war unter den jeweils gegebenen Umständen konsequentialistisch rational. Ein gewisses Problem bereitet höchstens die Reihenfolge der Stimmabgabe, da es sich hier nicht um ein reines Koordinationsproblem handelt (für die stärkere Gruppe gilt, daß sich am Ende nicht alle an der Wahl beteiligen müssen).

In Wirklichkeit ist den Bürgern der Stand der Stimmabgabe nicht bekannt. Sie erfahren ihn erst dann, wenn sie keine Entscheidung mehr treffen können. Sie sind gezwungen, sich zu einem Zeitpunkt zu entscheiden, in dem sie gerade der Information entbehren, die sie bräuchten, um die Konsequenzen ihres Handelns beurteilen zu können. Die Wahrscheinlichkeit dafür, daß sie in der Endabrechnung die entscheidende Stimme abgegeben haben, ist nun extrem gering. Aufgrund einer Veränderung der epistemischen Situation ist nun nicht mehr jede einzelne Stimmabgabe unter den jeweils gegebenen Umständen konsequentialistisch de facto optimierend, sondern i. allg. *keine einzige* Stimmabgabe. Allerdings gibt es nach wie vor jeweils mindestens eine Gruppe von Personen aus der Anhängerschaft von A resp. B, deren Stimmabgabe kollektiv konsequentialistisch optimierend ist. Es gibt *nur* eine Gruppe, wenn alle Bürger Anhänger entweder von A oder von B sind, wenn es also zu einem einstimmigen Ergebnis kommt. In allen anderen Fällen gibt es jeweils eine Menge von „ausschlaggebenden" Gruppen, die allein durch die Anzahl der Personen definiert ist: Wenn z. B. A mit einem Stimmenvorsprung von 8212 gewonnen hat, dann war die Stimmabgabe aller Gruppen aus der Anhängerschaft von A mit mehr als 8212 Mitgliedern kollektiv optimierend, während jede einzelne Stimme, aber auch die (kollektive) Stimmabgabe jeder Gruppe mit weniger als 8212 Mitgliedern nicht optimierend war. Diese Formulierung betrachtet die Kombinationen individueller Entscheidungen als Entscheidungen eines neuen Akteurs, eben der betreffenden Gruppe. Wer das aus einem konsequenten methodischen Individualismus ablehnt, kann nur die Stimmabgabe einer – mangels Unterscheidungsmöglichkeiten – beliebigen Gruppe von Personen als konsequentialistisch optimal bezeichnen; denn in jeder Gruppe, die mehr als 8212 Mitglieder umfaßt, ist wiederum die einzelne Entscheidung nicht konsequentialistisch optimal.

Die interpersonelle Übereinstimmung der Entscheidungssituationen entsteht erst durch eine drastische Informationsbeschränkung. Würde diese Informationsbeschränkung aufgehoben, könnte man die Handlungen der anderen Beteiligten zur Spezifikation der jeweiligen Entscheidungssituation heranziehen. Erst dann wäre die Entscheidungssituation wohlbestimmt. Der Konflikt zwischen individueller und kollektiver Optimierung ist Folge einer *Unterbestimmtheit der Entscheidungssituation* mangels relevanter Information. Die Bayes'sche Lösung mit Hilfe subjektiver Wahrscheinlichkeiten wäre jedoch unbefriedigend, weil die Freiheit des Handelns unter anderem darin zum Ausdruck kommt, daß die Zuordnung subjektiver Wahrscheinlichkeiten für eigenes und fremdes Handeln zu Paradoxien führt.

Bei *statischer* Betrachtung und unter der Voraussetzung *vollständiger Information* ist der Zusammenhang von individueller und kollektiver Bewertung im Rahmen des strikten Konsequentialismus *linear*. Linearität aber ist eine hinreichende Bedingung für die Vereinbarkeit von individueller

und kollektiver Maximierung. Individuelle und kollektive Bewertung sind bezüglich einer Bewertungsfunktion *linear*, wenn der Wert einer kollektiven Handlung h, die sich aus n individuellen Handlungen $h_1, h_2, h_3, \ldots, h_n$ zusammensetzt, gleich der *Summe der Werte der individuellen Handlungen* ist: $w(h) = \sum_{i=1}^{n} w(h_i)$.

Diese Linearität gilt auch für die Wahlsituation bei vollständiger Information. Zunächst scheint das wenig plausibel; denn die einzelnen Wahlakte der Anhängerschaft einer der beiden Parteien erhöhen den Wert der Gesamtsituation je individuell (und zwar um denjenigen Betrag, um den die Etablierung einer von A geleiteten Regierung höher zu bewerten ist als einer von B geleiteten, abzüglich der Nachteile des Wahlaktes selbst); aber am Ende ergibt sich nach k individuellen Wahlakten dennoch nur dieser Wert (abzüglich der Summe der Nachteile aller bis dahin getätigten Wahlakte), falls A siegt und falls B siegt, geht die Anhängerschaft von A auch noch dieses Vorteils verlustig. Die Summe der Werte der individuellen Wahlakte ist also bei weitem größer als der Wert der aggregierten individuellen Handlungen (der kollektiven Handlung der Wahlbeteiligung von k Bürgern). Diese naheliegende Analyse beruht jedoch auf einem Trugschluß. Auf die einzelnen Wahlakte der einen Gruppe folgen jeweils Wahlakte der anderen. Aus der Sicht der ersten Gruppierung bewirken diese Wahlakte jeweils ein Absenken des Wertes der Gesamtsituation – aus der Sicht der zweiten Gruppierung bewirken sie ein Ansteigen des Wertes der Gesamtsituation. Da wir hier jedoch die Konsistenz des strikten Konsequentialismus prüfen, kommen als strikt-konsequentialistische Akteure nur die Anhänger *einer* der beiden Gruppen in Betracht. Striktkonsequentialistische Rationalität ist ein Prädikat, das sich nicht auf individuelles Handeln isoliert anwenden läßt. Das je individuelle Handeln einer Gruppe von Personen ist dann strikt-konsequentialistisch rational, wenn es eine interpersonell invariante Bewertungsfunktion gibt, die von den Einzelhandlungen aller Akteure der Gruppe maximiert wird. Die Handlungen der anderen Gruppierung spielen daher für die Fragestellung, ob individuelle und kollektive Maximierung im Rahmen des strikten Konsequentialismus vereinbar ist, die Rolle *äußerer Umstände*. Diese „äußeren Umstände" senken allerdings den Wert der Gesamtsituation unmittelbar nach jeder Handlung wieder ab, so daß sich die Leistung derjenigen, die sich an der Wahl beteiligen, wie eine Sisyphusarbeit ausnimmt: Unabhängig davon, wie groß der Einsatz ist und wieviele sich am Ende beteiligen, entspricht der maximale Gewinn aller Wahlhandlungen in etwa demjenigen der einzelnen Wahlhandlungen (genaugenommen ist er sogar desto kleiner, je größer die Anzahl der Wahlbeteiligten der Gruppe am Ende ist). Diese Analyse zeigt, daß bei vollständiger Information die Linearität der *Wertaggregation nicht verletzt ist.*

Das Argument für die Vereinbarkeit von individueller und kollektiver Maximierung im Rahmen des strikten Konsequentialismus wurde jedoch

nur im objektivistischen Sinne entwickelt: Es wurde behauptet, daß es immer eine individuell maximierende Strategie gibt, die Bestandteil einer kollektiven maximierenden Strategie ist, aber nicht, daß diese Strategie auch dem betreffenden Akteur immer bekannt ist. Zwar steht eine solche optimale Strategie jedem Akteur einer idealen strikt-konsequentialistischen Gesellschaft in jeder Entscheidungssituation offen, aber da er sie in vielen Fällen nicht identifizieren kann, bleibt auch im Rahmen des strikten Konsequentialismus ein Koordinationsproblem bestehen, wenn die Information der Akteure unvollständig ist. *Es gibt im Rahmen des strikten ethischen Konsequentialismus bei statischer Betrachtung objektiv (im starken, nicht im schwachen Sinne, s. § 23) kein Koordinationsproblem.*

Das zweite Beispiel scheint jedoch für das Gegenteil zu sprechen. Der Professor hat kein Informationsproblem: Er weiß, in welchem Zustand der Rasen ist und welchen Einfluß die einmalige Überquerung haben würde. Jeder einzelne, der vor der Alternative steht, den Rasen zu überqueren, ist in der gleichen Situation: Er kennt ebenfalls den Zustand des Rasens, und er weiß, daß die einmalige Überquerung dem Rasen nicht zusätzlich schaden wird. Für viele Personen, die in dieser Situation sind, erscheint es konsequentialistisch rational, den Rasen zu überqueren. Es kann vermutet werden, daß die Anzahl der Personen, für die das zutrifft, ausreicht, um die Grünanlage zu ruinieren.

Wir müssen zwei Fälle unterscheiden: Im ersten Fall kommt es bei jeder Überquerung zwar zu einer marginalen Beschädigung des Rasens, was jedoch so wenig ins Gewicht fällt, daß die Bequemlichkeit der Abkürzung höher zu veranschlagen ist. In diesem Fall müßte zunächst geprüft werden, ob diese Annahme tatsächlich auch für eine strikt-konsequentialistische Wertfunktion plausibel ist – die Ausgangsthese gilt nur für den strikten Konsequentialismus. Dagegen spricht, daß die eigenen Interessen im Rahmen des strikten Konsequentialismus keine besondere Rolle spielen dürfen, da die zu maximierende Wertfunktion interpersonell invariant ist. Wenn man jedoch diese Annahme akzeptiert, ergibt sich eine paradoxe Konsequenz: Wenn tatsächlich der je individuelle Gewinn an Bequemlichkeit in der Wertfunktion höher rangiert als die je individuelle, wenn auch marginale Beschädigung des Rasens, dann ist nicht einzusehen, wie die kollektive Handlung, die sich aus einer Vielzahl von individuell konsequentialistisch rationalen Rasenüberquerungen zusammensetzt, unter ausschließlich diesen beiden Aspekten (der aggregierten Rasenbeschädigung und der aggregierten Bequemlichkeit) nicht konsequentialistisch rational sein kann. Es ist doch anzunehmen, daß der Gesamtgewinn an Bequemlichkeit dieser kollektiven Handlung der Summe individuellen Gewinns an Bequemlichkeit ebenso entspricht wie die Gesamtschädigung des Rasens als Folge dieser kollektiven Handlung der Summe der Einzelschädigungen. Der Einwand, erst ab einer bestimmten Anzahl von Überquerungen treten größere Beschädigungen durch die einzelne Überquerung auf,

entkräftet diese Aggregationsregel nicht; denn sie verlangt ja nicht, daß die Einzelschädingungen jeweils gleich groß sind, sondern daß die Gesamtschädigung der Summe der Einzelschädigungen entspricht. Selbst wer sich auf den nach traditionellem Naturverständnis (natura non facit saltus) wenig überzeugenden Standpunkt stellt, die einzelne Überquerung schade dem Rasen in keiner Weise, während die Überquerung von mehr als x Personen einen gewissen Schaden zufüge, kann damit die Linearitätsthese nicht widerlegen. Wenn neun Überquerungen keinen Schaden anrichten, dann ist es für die ersten neun Personen konsequentialistisch rational, den Rasen zu überqueren, aber der Zehnte darf den Rasen nicht mehr überqueren, wenn er weiß, daß er der Zehnte ist und seine Überquerung einen Schaden nach sich zieht, der von dem persönlichen Plus an Bequemlichkeit nicht aufgewogen wird. Wenn man davon ausgeht, daß der Zehnte i. allg. nicht weiß, daß er der Zehnte ist, dann liegt erneut ein Fall unvollständiger Information vor, der für die Ausgangsthese irrelevant ist. Eine ganz andere Situation ergibt sich, wenn wir annehmen, daß grundsätzlich keine einzelne Überquerung den Rasen schädigt, sondern erst die aggregierte Überquerung von z. B. zehn Personen. In diesem zweiten Fall verursacht auch der Zehnte, der den Rasen überquert, keinen Schaden, so paradox das klingen mag: Kein einzelner verursacht irgendeinen Schaden, sondern nur jeweils eine Gruppe von Personen. Um diesen Fall auch formal zu erfassen, müßte man auf die Präferenzschwellen-Theorie zurückgreifen. Erst in diesem zweiten Fall, der für das Rasenbeispiel empirisch sicher nicht gegeben ist, wäre die Linearität der Aggregation aufgehoben.

Das dritte Beispiel verführt eher wie das zweite dazu, die Linearitätseigenschaft anzuzweifeln, da die Belastung der Atmosphäre gleichzeitig erfolgt. Wenn man sich jedoch die Gesundheitsschädigung als eine stetige Funktion der getätigten Emissionen veranschaulicht, wäre ebenso Linearität der Aggregation garantiert wie bei reellwertiger Überschreitung gewisser gesundheitsschädlicher Grenzwerte.

§ 35 Eine Einschränkung der Vereinbarkeit

Es ist schwierig, ein plausibles Beispiel zu finden, in dem die Linearitätseigenschaft der Aggregation unter den beiden Annahmen der vollständigen Information (Objektivität) und der statischen Betrachtung auch bei sorgfältiger Analyse angezweifelt werden kann. Selbst die eindeutigen „Schwellenbeispiele", wie etwa die gemeinsame Unternehmung, ein Klavier von einem Zimmer in das nächste zu tragen, sind nur bei unvollständiger Information Illustrationen nicht-linearer Aggregation. Wenn eine Person P weiß, daß von den Anwesenden mehr als vier Personen bereit sind, das Klavier zu tragen (jeweils unter der Voraussetzung, daß weitere drei Personen ebenfalls dazu bereit sind), dann liegt ein reines Abstim-

mungsproblem vor und kein Koordinationsproblem im Sinne einer Verletzung der Kooperativitätsbedingungen durch strikt-konsequentialistisches Verhalten bei vollständiger Information[11].

Es gibt jedoch einen anderen Typ von „Schwellenbeispielen", der möglicherweise tatsächlich belegt, daß es auch bei vollständiger Information und statischer Betrachtung für strikte Konsequentialisten ein Koordinationsproblem gibt. Wenn wir die Wertfunktion eines (moralischen[12]) Vegetariers zugrunde legen, dann vermindert sich – ceteris paribus – der Wert eines Zustandes (monoton), wenn die Anzahl der zum menschlichen Verzehr geschlachteten Tiere steigt. Der einzelne Vegetarier steht jeweils vor der Entscheidung, ein einzelnes Fleischstück zu kaufen bzw. zu verzehren. Dieses Stück Fleisch stammt von einem schon geschlachteten Tier. Die Tötung dieses Tieres kann nicht dadurch rückgängig gemacht werden, daß eine einzelne Person auf den Verzehr des Fleisches verzichtet, aber auch die zukünftige Tötung des Schlachtviehs wird durch diesen Verzicht nicht beeinflußt. Es scheint keine kausale Beziehung zwischen der Anzahl der getöteten Tiere und dem je individuellen Verzicht auf ein bestimmtes Stück Fleisch zu bestehen. Selbstverständlich besteht jedoch eine kausale Beziehung zwischen dem Verzicht einer größeren Gruppe von Personen über einen längeren Zeitraum hinweg und der Anzahl der getöteten Tiere. Wenn man tatsächlich ausschließen kann, daß der je individuelle Verzicht *unter keinen Umständen* die Tötung eines Tieres verhindert, dann haben wir es hier mit einem besonderen Typ von Schwellenbeispiel zu tun. Danach gibt es keinen individuell entscheidenden Verzicht: Es ist keine Konstellation von Verzicht und Konsum möglich, in der der Verzicht auf *ein* Stück Fleisch genau die Schwelle überschreitet, die die Tötung eines zusätzlichen Tieres verhindert. Man muß diesen Fall deutlich von den anderen unterscheiden. Bisher mußte jeweils zugestanden werden, daß es bei vollständiger Information eine lineare Beziehung zwischen aggregiertem Wert und dem Wert der einzelnen Handlung gibt. Wenn die Annahmen dieses Beispieles jedoch zutreffen, dann gäbe es hier zum erstenmal *keine lineare Wertbeziehung zwischen individueller und kollektiver Handlung.*

[11] Wenn von zehn anwesenden konsequentialistisch rationalen Personen, sechs bereit sind, das Klavier zu tragen – was mit interpersonell invarianten Bewertungsfunktionen vereinbar ist –, dann könnte jeder einzelne verführt sein, seine Bereitschaft nach Kenntnis dieser Verteilung zurückzunehmen, da der Transport auch ohne seine Beteiligung sichergestellt ist. Wenn jeder sich in dieser Weise verhält, dann kommt es nicht zu dem – im Sinne der konsequentialistischen Bewertungsfunktion wünschenswerten – Transport. Für eine Gesellschaft von Konsequentialisten liegt diese Art „kindischen" Verhaltens tatsächlich nahe. Man darf jedoch nicht übersehen, daß dieses Verhalten nur unter der Annahme unvollständiger Information konsequentialistisch rational ist: Die Rücknahme der Bereitschaft, sich an dem Transport zu beteiligen, erfolgt ja nicht absolut zeitgleich, daher muß es eine Person geben, deren Entscheidung genau an der Schwelle liegt und den Transport definitiv vereitelt. In diesem Beispiel wäre es diejenige Person, die in chronologischer Reihenfolge als dritte ihre Bereitschaft zurückzieht.

[12] Das Attribut „moralisch" soll deutlich machen, daß die betreffende Person nicht aus geschmacklichen oder gesundheitlichen Gründen Vegetarier ist.

Eine in einem anderen Kontext entwickelte mathematische Theorie, die sich auf diese Form der Nicht-Linearität übertragen läßt[13], zeigt, daß der Wert einer kollektiven Handlung dann nicht mehr als Funktion der Werte der individuellen Handlungen darstellbar ist. *Konsequentialistisch rationales Handeln der einzelnen Personen einer Gruppe ist dann nicht mehr vereinbar mit einer kollektiv maximerenden Handlung:* Der moralische Vegetarier handelt im konsequentialistischen Sinne irrational. *Beide Kooperativitätsbedingungen sind bei Schwellenbeispielen dieser Art, d. h. bei Schwellenbeispielen mit nicht-linearer Aggregation, auch bei vollständiger Information und statischer Betrachtung nicht erfüllt.*

Es kann hier offenbleiben, ob es Fälle von nicht-linearer Aggregation tatsächlich gibt. Die kontraintuitiven Konsequenzen nicht-linearer Aggregation sprechen eher dagegen. So ist es bei nicht-linearer Aggregation ausgeschlossen, daß eine einzelne Handlung eine Auswirkung hat, obwohl eine Folge einzelner Handlungen Auswirkungen hat. Es ist daher ausgeschlossen, einen „Umschlagpunkt" anzugeben, bei dem die nächstfolgende Handlung eine Veränderung bewirkt. Obwohl es einen Umschlag gibt, erfolgt dieser Umschlag an keiner bestimmten Stelle. Es handelt sich dabei nicht um ein epistemisches Problem: Auch die allwissende Person könnte den Punkt des Umschlages nicht bestimmen. Die scheinbare Plausibilität der Präferenzschwellen-Theorie ergibt sich ebenso wie die These der nicht-linearen Aggregation aus einer Übertragung der Merkmale von Entscheidungssituationen mit unvollständiger Information auf die Bewertungsfunktionen selbst. Bei unvollständiger Information ist den Handelnden nicht bewußt, bei welcher Einzelhandlung der Umschlag erfolgt, und daher scheinen alle je individuell keinen Beitrag zu einer Veränderung zu leisten. Da es am Ende jedoch dennoch zu einer Veränderung kommt, vermutet man eine Wirkung, die sich nicht aus den Wirkungen der Einzelhandlungen zusammensetzt, sondern jeweils ausschließlich einer gewissen Anzahl von Einzelhandlungen zu verdanken ist – allerdings nicht in dem Sinne, daß die letzte der Einzelhandlungen, die die betreffende Anzahl ergibt, den Ausschlag gibt. Sollte es jedoch trotz dieser kontraintuitiven Implikationen gute Gründe dafür geben, auch Fälle nicht-linearer Aggregation zu berücksichtigen, dann wäre die Ausgangsthese leicht zu modifizieren:

Es gibt im Rahmen einer strikt-konsequentialistischen Theorie keine Interaktionssituation mit linearer Aggregation, in der (individuell) konsequentiali-

[13] Es handelt sich dabei um die sogenannte Präferenzschwellentheorie, die zunächst N. Georgescu-Roegen und später W. E. Armstrong entwickelt haben. Vgl. N. Georgescu-Roegen, „The Pure Theory of Consumer's Behavoir" *Quart. J. Econ.* 1 (1936) 545–593 sowie W. E. Armstrong, „The Determinateness of the Utility Function", *Econ. J.* 49 (1939). Dieser Ansatz wurde später ausgeweitet, um eine Methode der interpersonellen Vergleichbarkeit subjektiver Bewertungen zu entwickeln, vgl. L. A. Goodman/H. Markowitz, „Social Welfare Functions Based on Individual Rankings", *Am. J. Sociology* 18 (1952) 257–262 sowie J. Rothenberg, *The Measurement of Social Welfare*, Englewood Cliffs 1961.

stisch rationales Handeln und die (kollektive) Maximierung der Wertfunktion bei vollständiger Information und statischer Betrachtung unvereinbar sind.

§ 36 *Das dynamische Koordinationsproblem des strikten Konsequentialismus*

Das Handeln anderer Personen verändert die Rahmenbedingungen eigenen Handelns. Auch das eigene Handeln beeinflußt die Umstände späterer Entscheidungssituationen. Diese simple Tatsache hat gravierende Implikationen für die Analyse des Koordinationsproblems im Rahmen des strikten Konsequentialismus.

Wir sprechen von einer *dynamischen Betrachtung* des Problems, wenn diese Tatsache berücksichtigt wird – bleibt sie unberücksichtigt, handelt es sich um eine *statische Betrachtung*. Bisher haben wir uns auf eine statische Betrachtung beschränkt. Dabei wurde aufgezeigt, daß es bei vollständiger Information (und linearer Aggregation) kein Koordinationsproblem im Rahmen des strikten Konsequentialismus gibt. Allerdings ist dafür der strikte Konsequentialismus der einzige Kandidat. Für jede Abschwächung des strikten Konsequentialismus, etwa durch die Einführung einer individuellen Prärogative oder anderer individueller Prägungen der Wertfunktion, stellt sich das Koordinationsproblem ohnehin in voller Schärfe, da die Möglichkeit von Gefangenendilemma-Situationen ausschließlich durch strikt-konsequentialistische Wertfunktionen ausgeschlossen wird (vgl. dazu § 45). Das Koordinationsproblem scheint daher für den strikten Konsequentialismus weitgehend gelöst zu sein. Es stellt sich eher die Frage, ob diese radikale Methode der Lösung des Kooperationsproblems durch die Einführung einer einzigen interpersonell invarianten Wertfunktion, die von jeder rationalen Person maximiert wird, nicht einen zu hohen Preis darstellt. Die Kritik des Konsequentialismus unter dem Aspekt der Koordination ließe sich dann so zusammenfassen: Jede konsequentialistische, aber nicht strikt-konsequentialistische Theorie scheitert an dem generellen Koordinationsproblem der Gefangenendilemma-Situationen, und die strikt-konsequentialistischen Theorien sind als völlig unrealistische[14] normative Theorien ohnehin inadäquat. Überraschenderweise muß diese Kritik jedoch wesentlich verschärft werden: *Auch die strikt-konsequentialistischen Theorien scheitern am Koordinationsproblem – allerdings nur bei dynamischer Betrachtung.*

Das dynamische Koordinationsproblem läßt sich anhand einer Metapher erläutern: Angenommen, eine Stadt verfügt bisher nicht über Buslinien und macht sich daran, solche einzurichten. Sie wird diese Buslinien

[14] So z. B. J. L. Mackie, *Ethik. Auf der Suche nach dem Richtigen und dem Falschen*, Stuttgart 1981, Kap. 6.

nicht alle auf einmal schaffen können, sondern eine nach der anderen mit einem gewissen zeitlichen Abstand in Betrieb nehmen. Die Verwaltung könnte folgendermaßen vorgehen: Zunächst wird erwogen, welche Buslinie als erste (bei vorgegebener Gesamtlänge und Anzahl der Haltestellen) optimal wäre (welche Stadtviertel sie durchfahren sollte, wo die Haltestellen einzurichten wären etc.). Wenn der Zeitpunkt für die Einrichtung einer zweiten Buslinie gekommen ist, würde diese erneut optimal angelegt – jetzt natürlich unter Berücksichtigung der durch die erste Buslinie veränderten Rahmenbedingungen, u. s. w. Am Ende, so ist zu vermuten, ist das Gesamtsystem der Buslinien recht unübersichtlich und keineswegs optimal. Es wäre besser gewesen, wenn die Stadtverwaltung zunächst einen Plan des angestrebten Endzustandes gemacht hätte, um die einzelnen Buslinien dann so einzurichten, daß dieser Plan peu a peu erfüllt wird. Die einzelnen Buslinien stehen für individuelle Strategien. Das Optimierungsproblem der Buslinen entspricht dem Kooperationsproblem individueller Strategien, immer dann, wenn die einzelne Strategie die Rahmenbedingungen späterer Strategien verändert (so wie das Anlegen einer Buslinie die Rahmenbedingungen späterer Buslinien beeinflußt), führt die je individuelle Optimierung (selbst bei invarianter Wertfunktion) in der Aggregation im allgemeinen zu suboptimalen Ergebnissen. Die optimale Struktur ergibt sich nicht als eine Folge punktueller Optimierung.

Im allgemeinen läßt sich eine optimale Struktur diachronisch nur entwikkeln, wenn auf durchgängige punktuelle Optimierung verzichtet wird. Strukturelle und punktuelle Optimierung schließen sich i. allg. aus. Und das gilt, *ohne daß das Prinzip der linearen Aggregation infrage gestellt werden müßte!* Auch bei durchgängiger punktueller Optimierung im Sinne konsequentialistischer Rationalität hätte zu keinem Zeitpunkt eine (konsequentialistisch) bessere Entscheidung getroffen werden können. Trotz linearer Aggregation ist das Endergebnis schlechter als wenn die punktuellen Entscheidungen *strukturkonform* getroffen worden wären. Strukturkonformität heißt, das diachronische Vorgehen fiktiv durch ein synchronisches zu ersetzen, d. h. sich so zu verhalten, *als ob man nur eine Entscheidung zu treffen hätte.* In unserem Beispiel entspräche dem die Entscheidung für eine bestimmte Struktur der Buslinien (als Endzustand).

Der strikte ethische Konsequentialismus verlangt von der idealen moralischen Person punktuelle Optimierung. Punktuelle Optimierung aber sprengt i. allg. jede aggregative Struktur. Das Koordinationsproblem verlangt nach einem Verhalten, das sich an *struktureller Rationalität* orientiert. Nicht in allen Entscheidungssituationen ist diese Handlung in der Menge der je offenstehenden Handlungen eindeutig bestimmt. Das gilt ohnehin für Situationen mit unvollständiger Information, aber es gilt sogar für Situationen mit vollständiger Information, wenn sich die optimale Struktur durch eine Reihe unterschiedlicher Kombinationen individueller Handlungen erfüllen läßt. Die auf Entscheidungskoordination gerichtete

gesellschaftliche Kommunikation hat offensichtlich neben einer strategischen Rolle (der Beeinflussung subjektiver Bewertungen anderer Akteure) auch die Funktion, diese Unterbestimmtheit aufzuheben oder zumindest abzuschwächen.

Das dynamische Koordinationsproblem kann durch bloße punktuelle Rationalität auch bei invarianten Bewertungen nicht gelöst werden. *Strukturwahrende oder strukturschaffende Prozesse erfordern strukturell rationale Akteure. Bei jeder Wertfunktion gibt es jedoch Umstände, die eine punktuell optimierende Person strukturell irrational handeln läßt. Punktuelle Optimierung führt auch bei interpersonell invarianten Wertfunktionen, dynamisch betrachtet, zu strukturell irrationalen Ergebnissen.*

12. Kapitel: Strategische Verfälschung und Manipulation

In diesem Kapitel werden einige Ergebnisse der Logik kollektiver Entscheidungen, insbesondere das sog. Gibbard-Satterthwaite-Theorem, zu einer Kritik des Konsequentialismus herangezogen. Die Kritik richtet sich dabei gegen die konsequentialistische Rationalitätskonzeption als solche, aber auch gegen jede konsequentialistische Ethik mit Ausnahme des strikten ethischen Konsequentialismus. Die Grenze zwischen einer Analyse des Konsequentialismus als Rationalitätskonzeption und als ethischer Theorie ist damit weitgehend aufgehoben[1].

Konsequentialistische Ethiker können auf Einwände, die sich gegen die für den Konsequentialismus charakteristische Marginalisierung sittlicher Normen (vgl. § 27) richten, auf zweierlei Weise reagieren: Sie können diese Einwände für irrelevant erklären, da der „moralische Alltagsverstand" zur Beurteilung ethischer Theorien völlig ungeeignet sei[2], oder sie können auf bestehende empirische Bedingungen verweisen, die die kontraintuitiven Implikationen des ethischen Konsequentialismus ausschließen. Für diese zweite Reaktion spielt die *dispositionelle Strukturierung menschlichen Verhaltens* eine zentrale Rolle: Da der Mensch unfähig ist, von Fall zu Fall jeweils erneut eine die ganze Komplexität der Situation angemessen berücksichtigende Abwägung vorzunehmen, entwickelt er bestimmte Dispositionen. Die damit verbundene geringe Elastizität menschlichen Verhaltens läßt sittliche Normen, wie z. B. die Pflicht, ein Versprechen zu halten oder die Wahrheit zu sagen, auch wenn sie mit rigiden Verpflichtungsempfindungen verknüpft sein sollten, als konsequentialistisch rational erscheinen.

Es ist allerdings fraglich, ob die dispositionelle Strukturierung menschlichen Verhaltens tatsächlich hinreicht, um die angestrebte Versöhnung von Sittlichkeit und konsequentialistischer Theorie zu ermöglichen. Dagegen spricht die große Formbarkeit von Verhaltensdispositonen: Der Gewaltverbrecher kann ein liebender Familienvater sein, der Kriegsverbrecher ein seriöser und angesehener Bürger seiner Stadt. Wer seine Landsleute schätzt und rücksichtsvoll behandelt, kann zugleich von Fremdenhaß getrieben sein, wer auf Menschen Rücksicht nimmt, hat nicht notwen-

[1] Da die formalen Mittel für eine sorgfältige Darstellung der logischen Zusammenhänge relativ aufwendig sind und in L. Kern/J. Nida-Rümelin: *Logik kollektiver Entscheidungen*, München (Oldenbourg) 1994 (i. F. „LkE") in der gleichen Terminologie einschließlich der Beweise ausgeführt sind, beschränkt sich dieses Kapitel darauf, die Resultate für die Diskussion des Konsequentialismus fruchtbar zu machen.

[2] So besonders einige Vertreter des zeitgenössischen Utilitarismus, vgl. z. B. J. J. C. Smart, *An Outline of a System of Utilitarian Ethics*, Melbourne 1961, in revidierter Fassung als erster Teil von J. J. C. Smart/B. Williams: *Utilitarianism – For and Against*, Cambridge 1973 erneut publiziert.

digerweise die gleiche Verhaltensdisposition gegenüber Tieren[3]. Es gibt eine Art selektives Schuldgefühl, das nur bei moralischen Verfehlungen gegenüber nahestehenden Personen wirksam wird. Insbesondere gibt es das Phänomen einer recht differenzierten Anwendung von Kommunikationsnormen, wie die der Wahrhaftigkeit und des Vertrauens: Es hängt weitgehend von den Umständen, den Gesprächspartnern, den zu berücksichtigenden Interessen ab, in welchem Maße eine Person diese Verhaltensregeln befolgt. Wer sich dennoch relativ streng an bestimmte Verhaltensregeln hält, tut dies nicht notwendigerweise aus Gründen dispositioneller Prägung. Ein Verhalten dieser Art kann auch Ergebnis einer Folge von wohlabgewogenen Einzelentscheidungen sein. Dispositionelle Prägungen schränken die Entscheidungsfreiheit der handelnden Person ein, daher *ist eine schwach ausgeprägte dispositionelle Strukturierung des Verhaltens ein wesentliches Merkmal einer rationalen Person,* sofern Dispositionen nicht in einem behavioristischen Sinne verstanden werden.

Aber selbst wenn die empirische Psychologie bestätigen könnte, daß die dispositionelle Strukturierung menschlichen Verhaltens relativ starr sei, so daß sich die wichtigsten sittlichen Normen konsequentialistisch rechtfertigen ließen, wäre damit wenig gewonnen; denn eine ethische Theorie kann sich, ebenso wie naturwissenschaftliche Theorien, kontrafaktischen Tests nicht entziehen. So ist es eine legitime Frage, ob es gerechtfertigt sei, einen Teil der Menschheit zu versklaven, wenn dies nur optimale Konsequenzen im Sinne der jeweils zugrundegelegten Wertfunktion hätte (also etwa den Gesamtnutzen maximieren würde). Wenn die Auffassung zutrifft, daß eine Versklavung auch unter diesen speziellen Umständen nicht gerechtfertigt ist, dann wäre die betreffende konsequentialistische Theorie *auch dann widerlegt, wenn diese empirischen Bedingungen nicht auftreten* – das gilt sogar für den Fall, daß diese empirischen Bedingungen aus *naturgesetzlichen* Gründen ausgeschlossen sind. Daher ist es nicht nur zulässig, sondern notwendig, auch den ethischen Konsequentialismus ohne salvatorische Klauseln, die sich auf tatsächliche oder vermeintliche empirische Einschränkungen stützen, zu prüfen.

Wenn wir annehmen, daß es jeder konsequentialistisch rationalen Person möglich ist, punktuell zu optimieren, d.h. jede einzelne Entscheidung streng an den Kriterien des Konsequentialismus zu messen, dann ist auch die Entscheidung, in welcher Weise ich meine Mitmenschen informiere, Gegenstand konsequentialistischer Optimierung. Dies gilt generell für die Frage, unter welchen Bedingungen der rationale ethische Konsequentialist die Wahrheit sagt, aber speziell auch für die Frage, in welchem Um-

[3] Dies steht im Gegensatz zur Auffassung J. L. Mackies, daß eine auf der Kooperation „aktiv handelnder, vernünftiger Teilnehmer an einem zum Teil von Konkurrenzkampf bestimmten Leben" (S. 247) beruhende Ethik wegen der dispositionellen Strukturierung hinreiche, um Rücksichtnahme gegenüber Schwachen, Kranken, Minderheiten und sogar Tieren zu begründen. Vgl. J. L. Mackie, *Ethik*, Stuttgart 1981, Kap. 8, Abschn. 8.

fang der rationale ethische Konsequentialist seine eigenen Interessen und Wünsche strategisch verfälscht darstellen wird. Verfälschungen dieser Art sind Ausdruck einer „Strategieanfälligkeit" konsequentialistischer Rationalität, die unten genauer analysiert wird. Aber auch das bloße Arrangement der Entscheidungsfindung kann die Ergebnisse beeinflussen. Eine ausschließlich folgenorientierte Handlungsweise wird daher immer dann diese Arrangements beeinflussen, wenn damit eine Optimierung der Ergebnisse erreicht werden kann. Beeinflussungen der Arrangements der Entscheidungsfindung, um bestimmte gewünschte Ergebnisse zu erreichen, bezeichnen wir als „Manipulationen".

§ 37 Strategische Verfälschung

Im Gegensatz zu deontologischen Normensystemen empfiehlt der Konsequentialismus sowohl in seiner generellen Form wie als ethische Theorie eine *punktuelle Optimierung* der Folgen. Wenn eine Entscheidung optimale Folgen hat, dann ist sie auch dann rational bzw. moralisch geboten, wenn dabei bestimmte Normen verletzt werden. Dies gilt selbstverständlich auch für Entscheidungen, die die *Äußerungen* der handelnden Person betreffen (strategische Verfälschung der Präferenzen) oder für Entscheidungen, die die Arrangements der Entscheidungsfindung selbst zum Gegenstand haben (Manipulation).

Wie in § 20 und § 29 benötigen wir erneut einige begriffliche Hilfsmittel der Logik kollektiver Entscheidungen: Die betrachtete Gruppe von rationalen Konsequentialisten sei K, die Menge der zur Entscheidung anstehenden Alternativen X, die subjektiven Präferenzen der Konsequentialisten fassen wir in der Präferenzstruktur g zusammen. In jeder Entscheidungssituation kommt es aufgrund einer konsequentialistischen Abwägung der Beteiligten zu einer „Aggregation" von g, d.h. es erfolgt entweder explizit oder implizit in Gestalt individueller Entscheidungen eine Festlegung für eine der Alternativen aus X. Diese Festlegung kann auch als eine „kollektive Entscheidung" bezeichnet werden, denn sie bindet im Falle öffentlicher Entscheidungsverfahren die Mitglieder von K, und in den anderen Fällen einer impliziten Festlegung kommt sie aufgrund der einzelnen individuellen Entscheidungen zustande. Die Aggregation läßt sich formal durch eine Zuordnung darstellen, die einer gegebenen Präferenzstruktur eine („kollektive") Präferenzrelation oder zumindest eine kollektive Auswahlmenge zuordnet (LkE, Kap. 1). Einer Aggregationsregel entspricht eine Funktion, die beliebigen Präferenzstrukturen jeweils eine kollektive Präferenzrelation zuordnet. Da es in manchen Fällen ausreicht, jeweils diejenige (oder diejenigen) Alternative(n) aus X zu bestimmen, die aufgrund einer gegebenen Präferenzstruktur optimal ist (sind), läßt sich die Aggregationsregel auch als Funktion darstellen, die beliebigen Teilmengen von X jeweils eine Auswahlmenge zuordnet, also die

Menge derjenigen Elemente, die aufgrund der Aggregationsregel bei der gegebenen Präferenzstruktur optimal sind. Es gibt gute Gründe für die Forderung nach *eindeutigen* kollektiven Entscheidungen in dem Sinne, daß aufgrund einer Präferenzstruktur für beliebige Teilmengen S von X als Auswahlmengen nur Einermengen induziert werden[4]. Die Forderung nach einem eindeutigen Resultat bedeutet für kollektive Wohlfahrtsfunktionen, daß die kollektiven Präferenzrelationen strikte Ordnungen, also vollständig, asymmetrisch und transitiv sein müssen. Kollektive Wohlfahrtsfunktionen dieser Art werden strikte kollektive Wohlfahrtsfunktionen (SKWF) genannt.

DEF.: Eine Aggregationsregel (AR) f ist eine SKWF $:\leftrightarrow \forall g \in G$: $\forall X' \in \mathrm{Pot}(X)$, $\#X' \geq 3$: [f (g) ist vollständig, asymmetrisch und transitiv].

Nun kann aber auch der Definitionsbereich einer SKWF noch eingeschränkt werden, indem man fordert, daß die individuellen Präferenzrelationen, aus denen sich die Präferenzstrukturen zusammensetzen, ebenfalls strikte Ordnungen sein sollen.

Sei die Menge aller Präferenzstrukturen, die sich aus individuellen strikten Ordnungen zusammensetzen, mit G^S bezeichnet, so soll eine Aggregationsregel, die jeder Präferenzstruktur g aus G^S eine strikte Ordnung als kollektives Resultat zuordnet, eine *eingeschränkte strikte kollektive Wohlfahrtsfunktion* (S*KWF) genannt werden.

DEF.: Eine AR f* ist eine S*KWF $:\leftrightarrow \forall g \in G^S$: $\forall X' \in \mathrm{Pot}(X)$, $\#X' \geq 3$: [f* (g) ist vollständig, asymmetrisch und transitiv].

Da eine S*KWF nur Präferenzstrukturen aus strikten individuellen Ordnungen in strikte kollektive Ordnungen zu überführen gestattet, ist der Definitions- *und* Wertebereich einer S*KWF gegenüber dem Definitions- und Wertebereich einer KWF eingeschränkt, hingegen ist nur der Definitionsbereich einer S*KWF gegenüber einer SKWF eingeschränkt, nicht der Wertebereich.

Selbstverständlich gilt Arrows Theorem[5] auch für eingeschränkte strikte kollektive Wohlfahrtsfunktionen. Wir führen es im folgenden an, weil es zum Verständnis des Theorems der Strategieanfälligkeit notwendig ist. Dazu sind zunächst die Bedingungen von Arrow für eingeschränkte strikte kollektive Wohlfahrtsfunktionen umzuformulieren.

[4] In politischen Zusammenhängen würde eine Indifferenzklasse als kollektives Resultat keine Entscheidung für eine der Alternativen erlauben, so daß mangels Entscheidung der status quo selbst dann erhalten bliebe, wenn alle Entscheidungsbeteiligten irgendeine Alternative dem status quo vorziehen.

[5] K. Arrow, *Social Choice and Individual Values*, New York ²1963, Kap. V, in der hier gebrauchten Terminologie dargestellt und bewiesen (der Beweis von Arrow ist nicht korrekt) in *LkE*, Kap. 3.

Bedingung P:* (Pareto-Bedingung):

$\forall g \in G^S$: $\forall x,y \in X'$, $X' \in \mathrm{Pot}(X)$: $[\forall i \in K$: $\langle x,y\rangle \in \dot{g}(i) \rightarrow \langle x,y\rangle \in f^*(g)]$.

Bedingung I:* (Irrelevanz-Bedingung):

$\forall g,g' \in G^s$: $\forall x,y \in X', X' \in \mathrm{Pot}(X)$: $[\forall i \in K$: $[\langle x,y\rangle \in g(i) \leftrightarrow \langle x,y\rangle \in g'(i)]$ $\rightarrow$ $(\langle x,y\rangle \in f^*(g) \leftrightarrow \langle x,y\rangle \in f^*(g'))]$.

Bedingung D:* (Ausschluß der Diktatur)

$\neg\, \exists i \in K$: $\forall g \in G^S$: $\forall x,y \in X'$, $X' \in \mathrm{Pot}(X)$: $[f^*(g) = g(i)]$.

DEF.: Eine Person $i \in K$ mit der Eigenschaft
$\forall g \in G^S$: $\forall x,y \in X'$, $X' \in \mathrm{Pot}(X)$: $[f^*(g) = g(i)]$
ist ein *f*-Diktator.*

Mit diesen Bedingungen gilt die folgende Variante des Arrow-Theorems: *Erfüllt eine S*KWF f* die Bedingungen P* und I*, so kann sie nicht Bedingung D* erfüllen, d.h. dann gibt es einen f*-Diktator.*

Für kollektive Auswahlfunktionen bedeutet die Einschränkung auf eindeutige kollektive Resultate im obigen Sinne, daß sie für beliebige Teilmengen von X als Auswahlmengen Einermengen induzieren müssen. Kollektive Auswahlfunktionen dieser Art seien als *strikte kollektive Auswahlfunktionen* (SKAF) bezeichnet, wobei vorausgesetzt wird, daß ihnen Präferenzstrukturen zugrunde liegen, die aus individuellen Ordnungen zusammengesetzt sind.

DEF.: Eine Auswahlfunktion a ist eine strikte kollektive Auswahlfunktion SKAF :$\leftrightarrow$
$\forall g \in G$: $\forall X' \in \mathrm{Pot}(X)$: $[\# X' \geq 3 \rightarrow \# a_{f(g)}(X') = 1]$.

Analog zur weiteren Einschränkung strikter kollektiver Wohlfahrtsfunktionen kann auch für strikte kollektive Auswahlfunktionen gefordert werden, daß die zugrundeliegenden Präferenzstrukturen sich aus *strikten* individuellen Ordnungen zusammensetzen sollen. Solche kollektiven Auswahlfunktionen nennen wir *eingeschränkte strikte kollektive Auswahlfunktionen* (S*KAF).

DEF.: Eine Auswahlfunktion a* ist eine strikte kollektive Auswahlfunktion S*KAF :$\leftrightarrow$
$\forall g \in G^S$: $\forall X' \in \mathrm{Pot}(X)$: $[\# X' \geq 3 \rightarrow \# a^*_{f(g)}(X') = 1]$.

Die Bedingungen von K. Arrow für eingeschränkte strikte kollektive Auswahlfunktionen formuliert, lauten dann:

Bedingung der Pareto-Inklusivität P:*

$\forall g \in G^S$: $\forall X' \in \mathrm{Pot}(X)$, $X'' \subset X'$: $\forall x,y \in X'$, so daß $x \in X''$, $y \in X'$ und $y \notin X''$: $[\forall i \in K$: $\langle x,y\rangle \in \dot{g}(i) \rightarrow a^*_{f(g)}(X') \in X'']$.

Irrelevanzbedingung I:*

$\forall g,g' \in G^S$: $\forall x,y \in X'$, $X' \subset X$: $[\forall i \in K$: $[\langle x,y\rangle \in g(i) \leftrightarrow \langle x,y\rangle \in g'(i)] \rightarrow a^*_{f(g)}(X') = a^*_{f(g')}(X')]$.

Ausschluß der Diktatur D*:

$\neg \exists i \in K$: $\forall g \in G^S$: $\forall x \in X'$, $X' \subset X$, $x \neq a^*_{f(g)}(X')$:
$[\langle a^*_{f(g)}(X'),x\rangle \in \dot{g}(i)]$.

DEF.: Eine Person $i \in K$ mit der Eigenschaft
$\forall g \in G^S$: $\forall x,y \in X'$: $[X' \subset X \wedge a^*_{f(g)}(X') \neq x \rightarrow$
$\langle a^*_{f(g)}(X'),x\rangle \in \dot{g}(i)]$ ist ein a*-Diktator

Nun ist zu überprüfen, unter welchen Bedingungen Konsequentialisten auf eine „strategische" Veränderung ihrer Präferenzen verzichten würden. Zu diesem Zweck ist ein Postulat einzuführen, daß *die Strategieanfälligkeit, d.h. die vorteilhafte falsche Darstellung eigener Präferenzen ausschließen würde,* und zu prüfen, ob dieses Postulat von Konsequentialisten erfüllbar ist.

Zur Kennzeichnung von Veränderungen in Präferenzstrukturen führen wir die folgende Schreibweise ein:

$g/g'(i) = \langle g(i), \ldots, g(i-n), g'(i), g(i+1), \ldots, g(n)\rangle$

Sei g^S die Menge der zulässigen individuellen Präferenzrelationen, also der strikten individuellen Ordnungen, dann läßt sich Strategieanfälligkeit und Strategiefreiheit eingeschränkter strikter kollektiver Auswahlfunktionen wie folgt definieren.

DEF.: Eine Auswahlfunktion a* ist *strategieanfällig* bezüglich einer Präferenzstruktur $g \in G^S$:↔
$\exists g'(i) \in g^S$: $[\langle a^*_{f(g/g'(i))}(X'), a^*_{f(g)}(X')\rangle \in \dot{g}(i)]$.

DEF.: Eine Auswahlfunktion a* ist genau dann *strategiefrei,* wenn es keine Präferenzstruktur $g \in G^S$ gibt, bezüglich derer sie strategieanfällig ist.

Es gilt das folgende *Theorem* der Strategieanfälligkeit:

*Erfüllt eine eingeschränkte strikte kollektive Auswahlfunktion (S*KAF) a* die Bedingung der Strategiefreiheit S*, so kann sie nicht die Bedingung D* erfüllen, d.h. dann gibt es einen a*-Diktator*[6].

Die Beweisidee zu diesem Theorem beruht darauf, zunächst mit Hilfe zweier Lemmata zu zeigen, daß die Irrelevanz- und die Pareto-Bedingung

[6] Das Theorem der Strategieanfälligkeit wurde ursprünglich für „game forms" formuliert, d.h. für Funktionen, die jedes n-Tupel individueller Strategien (im Sinne der Spieltheorie) in ein endgültiges Resultat („outcome") überführen bzw. für „voting schemes", die jedem n-Tupel abgegebener Stimmen eine der Wahlalternativen zuordnen, s. A. Gibbard, „Manipulation and Voting Schemes: A General Result", *Econometrica* 41 (1973) 587–601, vgl. auch A. K. Sen, *Social Choice Theory,* in: *Handbook of Mathematical Economics,* Bd. III, hg. v. K. Arrow/M. Intriligator, Amsterdam 1986, S. 1130ff.

für die Strategiefreiheit einer Funktion a* unabdingbar sind. Sodann wird eine Konstruktion eingeführt, die es erlaubt, eine Funktion a* in eine Funktion f* zu überführen. Damit erfüllt auch f* die Irrelevanz- und die Pareto-Bedingung. Aufgrund von Arrows Theorem ist dann die Bedingung D^* nicht erfüllt, d.h. es gibt einen f*-Diktator. Da Bedingung D^* konstruktionsgemäß Bedingung D^* impliziert, ist dann erwiesen, daß eine strategische Auswahlfunktion a* die Bedingung D^* nicht erfüllt (LkE, Kap. 5)[7].

§ 38 Manipulation

Neben strategischer Verfälschung der Präferenzen ist die *Manipulationsanfälligkeit der Arrangements der Entscheidungsfindung* ein weiteres Problem konsequentialistischer Akteure. Diese Möglichkeit läßt sich nur ausschalten, wenn *Pfadunabhängigkeit* gegeben ist.

DEF.: Eine kollektive Auswahlfunktion ist *pfadunabhängig*[8] $:\leftrightarrow$
$\forall S,T \in \text{Pot}(X): [a(S \cup T) = a(a(S) \cup a(T))]$.

Aus dieser Definition ergibt sich unmittelbar, daß eine pfadunabhängige kollektive Auswahlfunktion manipulationsfrei sein muß, denn: wenn $a(a(S) \cup a(T))$, kann die Auswahl aus $S \cup T$ offenbar ein anderes Resultat zeitigen als die Auswahl aus $a(S) \cup a(T)$, was sich manipulativ nutzen läßt. Das ist jedoch nicht möglich, wenn $a(S \cup T) = a(a(S) \cup a(T))$.

Pfadunabhängigkeit sichert also Manipulationsfreiheit, daher ist es wichtig zu untersuchen, welche Eigenschaften pfadunabhängige kollektive Auswahlfunktionen haben. Die Pfadunabhängigkeit läßt sich wie folgt in Teileigenschaften zerlegen.

DEF.: Eine kollektive Auswahlfunktion ist *teilpfadunabhängig nach oben* (OPU) $:\leftrightarrow \forall S,T \in \text{Pot}(X): [a(S \cup T) \subseteq a(a(S) \cup a(T))]$.

DEF.: Eine kollektive Auswahlfunktion ist *teilpfadunabhängig nach unten* (UPU) $:\leftrightarrow \forall S,T \in \text{Pot}(X): [a(a(S) \cup a(T)) \subseteq a(S \cup T)]$.

Offensichtlich ist Pfadunabhängigkeit (PU) genau dann gegeben, wenn Teilpfadunabhängigkeit nach oben und nach unten gegeben ist: PU $\equiv$ UPU $\wedge$ OPU. Die obere Teilpfadunabhängigkeit OPU nun ist genau äquivalent zu α, der wichtigsten Auswahleigenschaft bei Mengenverringerung.

[7] Zum Beweis vgl. J.Green/J.-J. Laffont, *Incentives in Public Decision Making*, Amsterdam 1979, S. 14–19 sowie D. Schmeidler/H. Sonnenschein, „Two Proofs of the Gibbard-Satterwaite Theorem on the Possibility of a Strategy Proof Social Choice Function", in: *Decision Theory and Social Ethics*, H. W. Gottinger/W. Leinfellner, Dordrecht 1978, S. 227–234. Eine interessante Vereinfachung des Beweises hat A. Feldman, *Welfare Economics and Social Choice Theory*, Boston 1980, S. 206ff vorgeschlagen.

[8] Das Konzept der Pfadunabhängigkeit wurde von C. Plott, „Path Independence, Rationality, and Social Choice", *Econometrica* 41 (1973), S. 1075–1091 entwickelt.

α ist eine Forderung an die Auswahlfunktion. Eine Auswahlfunktion erfüllt α genau dann, wenn eine Alternative, die innerhalb einer Menge N ausgewählt wird, auch innerhalb einer Teilmenge von N ausgewählt wird, wenn die betreffende Alternative ebenfalls Element dieser Teilmenge ist.

Eigenschaft α*:*

$$\forall x \in X\colon \forall M,N \in \mathrm{Pot}(X)\colon [M \subset N \wedge x \in M \wedge x \in a_R(N) \rightarrow x \in a_R(M)]$$

Umgekehrt ist die untere Teilpfadunabhängigkeit UPU mit einer Auswahleigenschaft (ε) bei Mengenerweiterung äquivalent:

Eigenschaft ε*:* $\forall M,N \in \mathrm{Pot}(X)\colon [M \subset N \rightarrow \neg(a(M) \subseteq a(N))]$.

Wenn eine Auswahlfunktion die Eigenschaft ε hat, können bei der Auswahl aus der „größeren" Menge N (die M als Teilmenge enthält) nicht Elemente aus a(M) weggelassen und andere beibehalten werden, ohne in a(N) Elemente aufzunehmen, die nicht zu a(M) gehören.

Pfadunabhängigkeit ist daher äquivalent mit der Konjunktur der Bedingungen α *und* ε (vgl. LkE, Kap. 5, Lemma 3/5).

Auf dieses Ergebnis gestützt läßt sich beweisen, daß es in einer Gruppe von rationalen konsequentialistischen Akteuren generell (d.h. ohne Beschränkung der individuellen Präferenzen) Manipulationsfreiheit nur dann gibt, wenn das kollektive Resultat durch eine Veto-Gruppe bestimmt wird (LkE, Kap. 5).

§ 39 Schlußfolgerungen

In voller Allgemeinheit ist damit gezeigt, daß einer konsequentialistischen Gesellschaft im allgemeinen eine besondere Art der *Instabilität* eigen ist. Diese Instabilität ist Folge eines umfassenden optimierenden Verhaltens, das auch sekundäre Regeln, Verfahren der Entscheidungsfindung sowie Konventionen der Wahrhaftigkeit und des Vertrauens zum Gegenstand nimmt und daher zu Manipulationen der Arrangements der Entscheidungsfindung und zu strategischen Verfälschungen eigener Präferenzen führt. Sie ist in einer konsequentialistischen Gesellschaft nur zu beseitigen, wenn man unannehmbare Methoden der Entscheidungsfindung (z. B. daß eine Person allein für alle entscheidet – Verletzung der Bedingung D^* oder die Etablierung einer Entscheidungsoligarchie: „Veto-Gruppe") billigen würde. Dieses Ergebnis der Analyse spricht dafür, daß die universale Anwendung konsequentialistischer Entscheidungskriterien inakzeptabel ist. Die Ausklammerung bestimmter möglicher Entscheidungsalternativen aus dem Bereich konsequentialistischer Rationalität, wie sie sich z. B. in der Gestalt rechtlicher, konventionaler und sittlicher Normen äußert, ist Bedingung der Stabilität gesellschaftlicher Kooperation.

Allerdings gilt dieses Fazit nicht unter allen Bedingungen. Die Abstrak-

tion der Analyse hat zwar den Vorteil, die Strategie- und Manipulationsanfälligkeit einer Gesellschaft rationaler Konsequentialisten ohne jede weitere empirische Bedingung aufzeigen zu können, aber dies war nur um den Preis möglich, daß jeweils das gesamte Spektrum logisch möglicher subjektiver Bewertungen (Präferenzstrukturen) zugrunde gelegt wurde (dies erfolgte mit der Definition der Aggregationsregel, die *beliebigen* Präferenzstrukturen jeweils eine wohlbestimmte Auswahlfunktion bzw. kollektive Präferenzrelation zuordnet). Daher reicht diese allgemeine Analyse nicht hin, um etwa aufzuzeigen, daß eine Gesellschaft rationaler Konsequentialisten in jedem Falle, d. h. unabhängig von ihren spezifischen individuellen Wertfunktionen, instabil im Sinne der Strategie- und Manipulationsanfälligkeit sei, aber es reicht hin, um ein systematisches Defizit jeder konsequentialistischen Theorie deutlich zu machen; denn eine normative Theorie ist inadäquat, wenn sie für eine – realisierte oder hypothetische – Situation inakzeptable Konsequenzen hat. Die durch strategische Verfälschung der Präferenzen und Manipulation der Arrangements entstehende Instabilität einer Gesellschaft konsequentialistischer Akteure ist aber weder für die einzelnen Akteure wünschenswert, noch kann man sich eine plausible objektive Bewertung vorstellen, die diese befürworten würde.

Nur der strikte ethische Konsequentialismus scheint diesem Verdikt zu entgehen; denn eine Gesellschaft strikter rationaler ethischer Konsequentialisten zeichnet sich u. a. dadurch vor anderen konsequentialistischen Gesellschaften aus, daß ihre Mitglieder eine *übereinstimmende* Einschätzung der (moralischen) Rangfolge der Alternativen haben. Da wir immer den für den Konsequentialismus günstigsten Fall annehmen, sollten wir davon ausgehen, daß alle Personen in Übereinstimmung mit dieser moralischen Rangfolge der betreffenden Theorie des ethischen Konsequentialismus handeln – es gibt keine akrasia-Probleme oder epistemische Unstimmigkeiten. Unter diesen Bedingungen scheint strategische Verfälschung der Präferenzen und Manipulation der Entscheidungsarrangements nicht auftreten zu können. Erst in § 44 wird sich zeigen, daß diese Vermutung falsch ist, auch der strikte ethische Konsequentialismus hat ein Manipulations- und Strategieproblem, allerdings ein sehr spezielles.

13. Kapitel: Kritik zweiter Ordnung

§ 40 *Kritik zweiter Ordnung*

Die Disziplin der *Metaethik* steht zur Disziplin der normativen Ethik in einem ähnlichen Verhältnis wie die Wissenschaftsphilosophie zu den empirischen Einzelwissenschaften. So wie sich die Wissenschaftsphilosophie mit den in den Einzelwissenschaften zur Anwendung kommenden Begründungsverfahren kritisch auseinandersetzt und Modelle der Bestätigung oder Bewährung empirischer Theorien entwickelt, so befaßt sich die Metaethik mit dem Status und der Begründung wertender und normativer Urteile und der Bedeutung moralischer Ausdrücke.

Wenn wir im folgenden von „Kritik zweiter Ordnung" sprechen, dann ist damit ein spezielles Problem der Metaethik (in diesem weiten Sinne) angesprochen, das im ethischen Diskurs in jüngster Zeit zumindest implizit in der einen oder anderen Form eine Rolle gespielt hat. So rechtfertigt J. Rawls seine Theorie der Gerechtigkeit als Fairneß unter anderem auch mit dem „Stabilitätsargument"[1], wonach eine Gesellschaft, deren institutionelle Grundstruktur den beiden Gerechtigkeitsprinzipien entspricht, stabil sei, weil die Bürger einen Gerechtigkeitssinn entwickeln würden, der diese Institutionen vor Unterminierung bewahren würde. Umgekehrt argumentieren D. H. Hodgsons, B. Williams, E. E. Sleinis[2] und andere gegen eine utilitaristische Ethik mit dem Argument, ein Utilitarist könne nicht wünschen, daß eine Gesellschaft aus Utilitaristen besteht.

Zur Kritik zweiter Ordnung gehört die *Selbstanwendung einer Theorie:* Eine Theorie kann konsistent sein, aber den eigenen Kriterien nicht entsprechen. Nicht jede Theorie, die ihren eigenen Kriterien nicht entspricht, ist widersprüchlich. Die meisten Theorien lassen sich grundsätzlich nicht auf sich selbst anwenden, das gilt insbesondere für naturwissenschaftliche Theorien. Aber es sind nicht nur normativ-ethische Theorien, die für eine Kritik zweiter Ordnung in Frage kommen. Methodologische Theorien sind ebenfalls Kandidaten für eine Politik zweiter Ordnung. Max Webers Theorie der Wertfreiheit der Sozialwissenschaften ist ein prominentes Beispiel dafür, denn sie scheint den eigenen Kriterien der Wissenschaftlichkeit nicht zu entsprechen[3]. Aus der Logik wissen wir, welche philosophischen Probleme die Selbstanwendung von Aussagen aufwirft. Daher ist es notwendig, sich zunächst klarzumachen, welcher Stellenwert der Kritik zweiter Ordnung in der Ethik zukommt und in welchem Sinne man von der Selbstanwendung einer normativ-ethischen Theorie sprechen kann.

[1] Vgl. J. Rawls, *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975, Kap. 8.

[2] D. H. Hodgson, *Consequences of Utilitarianism*, Oxford 1967, Kap. 2; B. Williams, *Kritik des Utilitarismus*, Kap. 6; E. E. Sleinis, „A Problem in Utilitarianism", *Journal of Value Inquiry* 16 (1982) 75–77.

[3] Vgl. Verf., „Moderne Wissenschaftskonzeptionen", in: *Kulturelle Konfrontation oder interkulturelles Lernen*, hg. von der Otto Benecke Stiftung, Baden-Baden 1987, S. 81–120.

Zum besseren Verständnis sei zunächst ein Beispiel einer Kritik zweiter Ordnung, auf eine *deskriptive* Theorie angewandt, betrachtet. Kardinal Bellarmin hatte gegen Galileis Umstürzung der Astronomie nicht nur „wissenschaftsinterne" Argumente, sondern auch die These vorgebracht, es sei um den Seelenfrieden der Menschen schlecht bestellt, wenn sie ihren Glauben an die im Weltmittelpunkt ruhende Erde verlören[4]. P. Feyerabend schließt sich diesem Typus der Argumentation ausdrücklich an: Er hält ihn für legitim, auch wenn er andere Wertungen zugrunde legen würde. Der erkenntnistheoretische Anarchist verteidigt oder bekämpft eine Theorie je nachdem, ob er ihre voraussichtlichen gesellschaftlichen Auswirkungen begrüßt oder verabscheut[5]. Ist eine Kritik zweiter Ordnung in dieser Form zulässig? Offensichtlich muß zumindest als eine methodologische Voraussetzung für die Zulässigkeit dieser Art der Kritik zweiter Ordnung gelten, daß es für die in Frage stehende Theorie keine unabhängigen Prüfungskriterien gibt. Wer als erkenntnistheoretischer Realist dagegen davon ausgeht, daß eine Theorie Ausschnitte der Realität mehr oder weniger zutreffend beschreibt oder modellhaft-idealisierend wiedergibt, der kann diesen Typus einer Kritik zweiter Ordnung, der die praktischen Folgen einer Theorie zum Kriterium erhebt, nicht akzeptieren. Es kann hier offen bleiben, ob gleiches auch für andere erkenntnistheoretische Positionen gilt.

§ 41 Ethik und Entscheidungsfindung

Angenommen, es ließe sich folgender Sachverhalt aufzeigen: Es gibt keine konsequentialistische ethische Theorie (T), die die Eigenschaft hat, daß eine Gesellschaft von idealen moralischen Personen P_T die Wertfunktion dieser Theorie (W_T) maximiert[6]. Wäre mit dem Nachweis dieses Sachverhaltes der Konsequentialismus widerlegt?

In der Diskussion um den Utilitarismus ist in verschiedenen Varianten die These begründet worden, ein Utilitarist könne sich unter bestimmten Bedingungen nicht wünschen, daß alle Personen einer Gesellschaft Utilitaristen sind. Die Proponenten dieser These sind meist der Auffassung, damit sei der Utilitarismus widerlegt[7]. Diese Schlußfolgerung ist jedoch keineswegs selbstverständlich. An dem oben erwähnten Beispiel der Galilei-Kritik Bellarmins wird deutlich, daß der praktische Nutzen oder Schaden einer Theorie i. allg. kein Argument für ihre Richtigkeit bzw. Falschheit darstellt.

[4] Vgl. J. Broderick, *Robert Bellarmin. Saint an Scholar*, London 1961.

[5] Vgl. P. Feyerabend, *Wider den Methodenzwang*, Frankfurt/M. 1976, S. 262–270.

[6] Diese Annahme entspräche der These: Keine konsequentialistische Theorie erfüllt die starke Kooperativitätsbedingung aus § 31.

[7] Vgl. als ein Beispiel für viele: E. E. Sleinis, „A Problem in Utilitarianism", *Journal of Value Inquiry* 16 (1982) 75–77. Ein ähnliches Argument gebraucht schon G. E. Moore, *Principia Ethica*, Stuttgart 1970, S. 218ff.

Wenn auch für ethische Theorien gilt, daß die Kriterien ihrer Adäquanz unabhängig von der praktischen Anwendung dieser Theorien sind, dann ist eine ethische Auffassung keine bloße Frage der *Entscheidung* (so wie auch Galilei sich nicht entscheiden konnte, daran zu glauben, daß die Erde ruht). Wer (subjektive) Gründe hat, eine Aussage oder eine Theorie für wahr zu halten, der kann sich i. allg. nicht mehr *entscheiden*, diese Aussage oder diese Theorie für falsch zu halten – auch dann nicht, wenn er unterdessen weiß, daß diese epistemische Einstellung schädlich ist (sei es für ihn persönlich oder auch unter moralischen Aspekten). Aber selbst wenn er sich entscheiden könnte, etwa weil ihm ein Mittel (z. B. ein Medikament oder „Gehirnwäsche") zur Verfügung steht, seine Annahmen zu beeinflussen, würde es nicht gegen die *Richtigkeit* der Annahmen sprechen, daß er sich aus Gründen eines praktischen Vorteils gegen sie entscheiden würde. Ebenso scheint für den Utilitarismus zu gelten: Selbst wenn sich der Utilitarist wünscht, daß er selbst, bestimmte Personen oder alle Mitglieder der Gesellschaft unter gewissen Bedingungen keine Utilitaristen sind, und dieser Wunsch rein moralisch – also utilitaristisch – motiviert ist, könnte der Utilitarismus dennoch eine adäquate ethische Theorie darstellen.

Das Problem läßt sich zuspitzen: Ist die ideale moralische Person des Utilitarismus – wir bleiben zunächst bei diesem besonders eindeutigen Beispiel einer ethischen Theorie – notwendigerweise ein Utilitarist? Allgemeiner formuliert: Ist die ideale moralische Person einer Theorie Proponent der Theorie? Es ist nicht selbstverständlich, daß diese Frage mit „Ja" zu beantworten ist. So enthält die Ballistik wesentliche Elemente einer Theorie des idealen Weitwurfes (Abwurfwinkel in Abhängigkeit von der Anfangsgeschwindigkeit, der Größe und dem Gewicht des Wurfgegenstandes etc.), ohne daß daraus gefolgert werden kann, der ideale Leichtathlet müsse Ballistik beherrschen. Möglicherweise würden ihn theoretische Kenntnisse dieser Art sogar eher behindern. Moralisch richtiges Handeln setzt keine theoretischen Kenntnisse voraus, und so scheint es, daß die ideale moralische Person einer Theorie nicht Proponent dieser Theorie sein muß.

Wir müssen offensichtlich zwei Fälle unterscheiden. Der erste Fall ist gegeben, wenn eine ethische Theorie ein Kriterium richtigen Handelns formuliert, ohne zugleich *eine Theorie der richtigen Entscheidungsfindung* zu sein. Eine Theorie der richtigen Entscheidungsfindung gibt an, in welcher Weise die ideale moralische Person ihre Entscheidungen trifft. Der zweite Fall wäre gegeben, wenn eine ethische Theorie *nicht nur ein Kriterium richtigen Handelns entwickelt*, sondern zugleich (zumindest implizit) bestimmt, in welcher Weise die ideale moralische Person ihre Entscheidungen trifft.

Im ersten Fall wäre das Argument der oben erwähnten Utilitarismus-Kritiker ein non sequitur: Wenn der Utilitarismus nicht als *Theorie der*

richtigen Entscheidungsfindung zu verstehen ist, sondern ausschließlich als *handlungskriteriale* Theorie, dann ist die Frage, ob Utilitaristen das Handlungskriterium des Utilitarismus je individuell oder kollektiv erfüllen, logisch unabhängig von der Frage, ob der Utilitarismus eine zutreffende Theorie ist.

Wenn der Utilitarismus jedoch auch als eine Theorie der Entscheidungsfindung (prozedural) verstanden wird (im zweiten Falle also), ist es zumindest naheliegend anzunehmen, daß die ideale moralische Person des Utilitarismus ein Utilitarist ist[8]. Im zweiten Fall würde der Utilitarismus die (schwache) Kooperativitätsbedingung nicht erfüllen.

Aus den Aussagen:

1) Eine adäquate (konsequentialistische) Ethik erfüllt die (schwache) Kooperativitätsbedingung;

2) eine gute utilitaristische Theorie ist keine bloße handlungskriteriale Theorie, sondern enthält zugleich eine (prozedurale) Theorie der moralischen Entscheidungsfindung;

3) die ideale moralische Person einer utilitaristischen Theorie ist ein Utilitarist;

4) der ideale Utilitarist trifft Entscheidungen, für die gilt, daß sie
a) individuell die Wertfunktion der utilitaristischen Theorie nicht maximieren oder
b) kollektiv die Wertfunktion der utilitaristischen Theorie nicht maximieren,

folgt:
Der Utilitarismus ist keine adäquate (konsequentialistische) Theorie.

Dieses Ergebnis läßt sich problemlos auf den Konsequentialismus generell übertragen: Wenn es keine konsequentialistische ethische Theorie (T) gibt, die die Eigenschaften hat, daß eine Gesellschaft von idealen moralischen Personen P_T die Wertfunktion dieser Theorie W_T maximiert, dann

[8] Die hier angesprochene Problematik hat die ethische Diskussion seit ihren Anfängen begleitet. Platon kann dabei als extremer Verfechter der *theoretischen* Fundierung sittlicher Entscheidungen angesehen werden (eine intellektualistische Position, wie sie besonders im Gorgias-Dialog deutlich wird), während Aristoteles' Figur des spoudaios für eine nicht-intellektualistische Interpretation zumindest offen ist, wenn sie eine solche auch nicht erzwingt. Bestärkt wird diese nicht-intellektualistische, sittliche Interpretation im zehnten Buch der *Nikomachischen Ethik* (NE X,6-9), wo Aristoteles betont, daß theoretische Kenntnisse in keiner Weise eine besondere sittliche Qualifikation mit sich bringen. Ein besonders pointierter Vertreter dieses sittlichen Modells der Aristoteles-Interpretation ist: G. Teichmüller: *Die praktische Vernunft bei Aristoteles*, Gotha 1879; dagegen steht das theoretische Modell der Aristoteles-Interpretation z. B. von J. Walter, *Die Lehre von der praktischen Vernunft in der griechischen Philosophie*, Jena 1874. Zum Begriff des spoudaios vgl. auch die eigenwillige Interpretation von E. Voegelin, *Die Neue Wissenschaft der Politik*, München 1959, Kap. II.5, sowie *Anmamnesis*, München 1966, S. 124–133. H. Kuhn hat das Fehlen einer theoretischen Fundierung sittlicher Normen bei Aristoteles überzeugend kritisiert: „Wissenschaft der Praxis und praktische Wissenschaft", in: *Werden und Handeln* (Festschrift für Frhr. v. Gebsattel), Stuttgart 1963, S. 157–190.

wäre aufgezeigt, daß der Konsequentialismus eine inadäquate ethische Theorie ist. Die Ausgangsfrage dieses Abschnittes kann insoweit bejaht werden. Die genauere Analyse hat jedoch gezeigt, daß die Art und Weise einer potentiellen Widerlegung davon abhängt, ob der Konsequentialismus eine prozedurale Komponente hat und ggf. welche Eigenschaften diese prozedurale Komponente aufweist.

§ 42 *Konstruktivismus und Kritik zweiter Ordnung*

Es wurde schon ausgeführt, daß jede adäquate ethische Theorie (explizit oder zumindest implizit) eine Theorie der idealen moralischen Person enthält. Eine *konstruktive ethische Theorie* ist dadurch gekennzeichnet, daß sie ihre moralische Konzeption als diejenige Handlungsorientierung charakterisiert, die eine vernünftige moralische Person prägen und befürworten würde.

Die *kontraktualistische Variante des Konstruktivismus*[9] charakterisiert ihre ethische Konzeption durch ein Verfahren der Entscheidung vernünftiger Personen für ein Normensystem, oder allgemeiner: eine Handlungsorientierung, von der sie wünschen, daß sie ihr Verhalten in Zukunft leiten sollte. Eine kontraktualistisch begründete Gerechtigkeitskonzeption ist keine (ausschließlich) kriteriale Konzeption in dem oben eingeführten Sinne, sondern hat als eine Bedingung ihrer Adäquanz, daß sie als Handlungsorientierung (u. U. nur innerhalb der betreffenden gesellschaftlichen Rahmenbedingungen ihres Anwendungsbereiches) tauglich ist – sie enthält daher immer neben einem kriterialen auch ein prozedurales Element. In kontraktualistischen Ethiken wird das Moment der Konstruktion explizit: Die betroffenen Personen schaffen das für sie verbindliche Normensystem: Damit ist nicht notwendigerweise eine subjektivistische Metaethik verbunden. Die Kriterien der Wahl und die Charakteristika der Wahlsituation können normativ gebunden bleiben[10].

Der *Kantische Konstruktivismus* verlangt, daß eine Konzeption des moralisch Richtigen zugleich Grundlage politischer Entscheidungen in der Öffentlichkeit ist[11]. Sie formuliert nicht nur ein Kriterium für die Entscheidungen des Gesetzgebers (der Vertrag als eine „bloße Idee der Vernunft, die aber ihre unbezweifelte praktische Realität hat: nämlich den Gesetzgeber zu verbinden, daß er seine Gesetze so gebe, als sie aus dem vereinten

[9] J. Rawls hat diese Variante zur Interpretation seiner Theorie der Gerechtigkeit in „Kantian Constructivism in Moral Theory", *Journal of Philosophy* 77 (1980) 515–572, dt. in: *Die Idee des politischen Liberalismus*, Frankfurt/M. 1992, Kap. 2 entwickelt.

[10] Unter den klassischen kontraktualistischen Theoretikern sticht nur T. Hobbes mit einer rein empirischen Charakterisierung der Wahlsituation hervor, während die Konstruktion des Normensystems bei J. Locke und in seiner Nachfolge auch bei R. Nozick explizit normativen Beschränkungen unterliegt. Auch J. Rawls wäre mißverstanden, wenn man die dort entwickelte Konzeption des Urzustandes („original position") als außermoralisch interpretieren würde.

[11] Daraus erklärt sich auch das Publizitätskriterium der Friedensschrift, vgl. I. Kant, *Zum ewigen Frieden* (1795) in: Gesammelte Schriften (Akademiausg.) Bd. VIII, (Repr. 1968).

Willen eines ganzen Volkes haben entspringen können"[12], sondern ist zugleich ein „praktisches Prinzip" als „Folge einer Willensbestimmung a priori"[13] jedes einzelnen.

Man könnte sagen, konstruktivistische Ethiken nehmen die Menschen, auf die sie sich beziehen, als *moralische Personen* ernst. Dies wird deutlich, wenn man die Argumentation nicht-konstruktivistischer Ethiker zum Vergleich heranzieht. J. J. C. Smart[14] betrachtet den Fall, in dem sich ein Utilitarist in einer von magischen Vorstellungen geprägten Gesellschaft befindet, deren Normen auf archaischen Tabus beruhen. Smart hält es für möglich, daß die Aufrechterhaltung der Tabu-Ethik unter einem utilitaristischen Aspekt besser sein könnte als die „moralische Anarchie", die sich einstellen würde, wenn die Mitglieder dieser Gesellschaft ihren Glauben an magische Kräfte verlieren. Ein Utilitarist würde in diesem Falle nicht für die Übernahme einer utilitaristischen Einstellung plädieren und damit zu einer Schwächung der Tabu-Ethik beitragen. Eine Verallgemeinerung dieses Argumentes auf die Anwendungsbedingungen normativer Ethik generell hieße, alle Menschen nicht als moralische Personen anzusehen; denn für den Smartschen Utilitaristen sind die Mitglieder für das, was sie tun, kaum verantwortlich zu machen. Die Vorwürfe, die Smart einer der utilitaristischen Ethik zugänglichen Person machen kann, wenn sie etwas tut, was dem Gemeinwohl abträglich ist, sind auf die Mitglieder der Tabu-Gesellschaft nicht anwendbar. Er könnte ihnen ggf. nur Abweichungen von den tradierten Normen vorhalten, aber diese Vorhaltungen wären unehrlich. Schließlich richtet sich Smarts Kritik in Wahrheit nicht gegen die Abweichung von den Regeln der Tabu-Ethik, sondern gegen die Minderung der in dieser Gesellschaft verwirklichten Nutzensumme.

Damit ist der wichtigste Grund genannt, der für eine konstruktivistische Ethik spricht: Die Adressaten normativer Ethik sind moralische Personen, daher sollte eine adäquate ethische Theorie ihre moralische Konzeption als diejenige Orientierung charakterisieren, die eine vernünftige moralische Person idealiter leiten würde.

§ 43 Zerstörung normativer Institutionen

In diesem Zusammenhang müssen wir ein Argument aufgreifen, das wir schon in § 27 unter einem anderen Aspekt angesprochen hatten: das der Unvereinbarkeit von Konsequentialismus und normativen Institutionen.

Verhalten im Einklang mit normativen Institutionen, wie z. B. der des Versprechens, der Wahrhaftigkeit und des Vertrauens, wird durch eine Vielzahl von Sanktionen gestützt. Nur ein kleiner Teil dieser Sanktionen

[12] *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis* (1793), in: Gesammelte Schriften (Akademieausg.) Bd. VIII, (Repr. 1968), S. 250.

[13] Vgl. I. Kant, *Kritik der praktischen Vernunft* (1788), in: *Gesammelte Schriften* (Akademieausg.), Bd. V, (Repr. 1968), 1. Teil, I. Buch, 2. Hauptstück.

[14] J. J. C. Smart/B. Williams, *Utilitarianism – For and Against*, Cambridge 1963.

ist rechtlich gesichert, ein anderer Teil schlägt sich in gesellschaftlicher Verachtung, Entzug von Zuneigung und Nachlassen der Kooperationsbereitschaft nieder. Auch Gewissensbisse und andere Nachteile in foro interno kann man als einen besonderen Typ von Sanktionen betrachten[15]. Es ist naheliegend anzunehmen, daß normative Institutionen aber auch in hohem Maße auf dem Gerechtigkeitssinn und anderen Elementen der Sittlichkeit beruhen. Ohne eine Handlungsorientierung, für die Verhalten im Einklang mit bestimmten normativen Institutionen als moralisch (intrinsisch) wertvoll gilt, würden viele dieser normativen Institutionen auf Dauer nicht bestehen bleiben. Je nach der Wertfunktion einer konsequentialistischen Theorie mag eine solche Entwicklung im Einzelfall sogar zu begrüßen sein; denn möglicherweise ist die Fixierung vieler Menschen auf bestimmte rigide normative Institutionen einer Optimierung dieser Wertfunktion abträglich. Aber zur Kritik des Konsequentialismus reicht es hin, einige oder sogar nur eine normative Institution zu nennen, deren Aufrechterhaltung von jeder adäquaten konsequentialistischen Wertfunktion befürwortet wird, die aber durch die Dynamik der Interaktion konsequentialistischer Akteure zerstört würde. Wir betrachten dabei den Fall des strikten ethischen Konsequentialismus als derjenigen Spielart, die gegen dieses Argument am besten gefeit zu sein scheint, und ergänzen damit zugleich die Analyse der Koordinations- und Verfälschungsproblematik des Konsequentialismus der §§ 31–36 und 37–39.

T sei eine strikt konsequentialistische ethische Theorie mit der idealen moralischen Person P_T und der Wertfunktion W_T. Aufgrund der im letzten Abschnitt ausgeführten Überlegung sei T darüber hinaus konstruktivistisch. Daher stimmt nicht nur das Handeln von P_T mit T überein, insofern es je individuell und punktuell W_T optimiert, sondern darüber hinaus ist P_T eine Person, deren moralische Orientierung T entspricht: Der Komplex der handlungsleitenden (vorausgehenden und motivierenden Absichten) erfüllt die Bedingungen von T: P_T ist T-Konsequentialist.

Wir betrachten eine Gesellschaft, die aus in dieser Weise charakterisierten idealen moralischen Personen besteht. Diese Personen werden die Regeln einer normativen Institution in der jeweiligen Entscheidungssituation genau dann einhalten, wenn sie dadurch die Wertfunktion W_T (punktuell) maximieren. Die normative Institution ist dabei für die Entscheidung des Konsequentialisten nur insoweit relevant, als das Bestehen der normativen Institution möglicherweise einen kausalen Einfluß auf die Folgewirkungen der Handlung hat. Ein T-Konsequentialist wird ein gegebenes Versprechen genau dann einhalten, wenn die Einhaltung dieses Verspre-

[15] Zwichen E. Tugendhat und U. Wolf hat es jüngst eine interessante Kontroverse um die Frage gegeben, ob sich Moralität – wie Tugendhat meint – auf sanktionierte moralische Forderungen beschränkt oder nicht. Vgl. E. Tugendhat, *Probleme der Ethik*, Stuttgart 1984, insbes. die ersten drei Vorlesungen u. Retraktationen, und U. Wolf, *Das Problem des moralischen Sollens*, Berlin/New York 1984.

chens in der konkreten Entscheidungssituation die Wertfunktin W_T maximiert.

Die normative Institution des Versprechens ist eine in allen menschlichen Gesellschaften fundamentale Regel der Handlungskoordination. Es ist daher anzunehmen, daß der in einer Gesellschaft insgesamt verwirklichte und durch W_T repräsentierte Wert abnimmt, wenn sich die Institution des Versprechens nicht aufrechterhalten ließe.

Die normative Institution des Versprechens ist nur solange wirksam, als eine Person (P), der etwas (h) versprochen wurde, a) mit großer Wahrscheinlichkeit erwarten kann, daß die Person (P*), die ihr das Versprechen gegeben hat, h vollzieht, und b) mit größerer Wahrscheinlichkeit erwarten kann, daß P* h vollzieht, als wenn P* ihr Versprchen nicht gegeben hätte[16].

Der rationale Konsequentialist fühlt sich an eine Regel moralisch nicht gebunden. Das Einhalten einer Regel hat für sich genommen keinerlei moralischen Wert. Im strengen Sinne gibt es daher für den rationalen Konsequentialisten keinerlei moralische Verpflichtung gegenüber P als Folge des gegebenen Versprechens; denn alle Verpflichtungsurteile sind ausschließlich Implikationen des Optimierungskriteriums. Da P als rationaler Konsequentialist ebenfalls keine moralischen Rechte aus dem gegebenen Versprechen als solchem ableiten wird (sondern höchstens indirekt, falls das Versprechen einen kausalen Einfluß auf die Folgewirkungen der offenstehenden Handlungen hat), verliert unter diesen Voraussetzungen ein von Utilitaristen häufig gebrauchtes Argument seine Gültigkeit: Die Enttäuschung Ps darüber, daß P* sein Versprechen nicht eingehalten hat, kann für die konsequentialistische Folgenabwägung ebensowenig eine Rolle spielen wie das schlechte Gewissen und andere Sanktionen in foro interno für P*. Auch die zukünftige Erwartungshaltung gegenüber P* von seiten Ps dürfte keine Rolle spielen, da die konsequentialistische Orientierung öffentlich ist – Ps Erwartungen gegenüber P* können sich nur dann ändern, wenn P P* irrationales Verhalten unterstellt. P konnte auch vor dem Versprechensbruch nicht erwarten, daß P* seine Versprechen aufgrund eines Verpflichtungsgefühls gegenüber der normativen Institution des Versprechens einhält, da P wußte, daß P* ein rationaler Konsequentialist ist. Darüber hinaus wußte P, daß P* wußte, daß P wußte, daß P* ein rationaler Konsequentialist ist, und daher wußte P, daß P* nicht befürchten mußte, P zu enttäuschen (diese Enttäuschung hätte sich sonst in der konsequentialistischen Abwägung P*s zu Ungunsten eines Versprechensbruches niederschlagen können).

Es ist daher nicht die Abwägung der kausalen Folgen einer Handlung, die normative Institutionen wie die des Versprechens in einer Gesellschaft rationaler Konsequentialisten zerstört, sondern es ist die Veränderung der

[16] Vgl. dazu D. H. Hodgson, *Consequences of Utilitarianism*, Oxford 1967, Kap. 2.

kausalen Folgen der Handlungen, die dieses Ergebnis hervorbringt. In einer von normativen Institutionen stark geprägten Gesellschaft ist es für einen Konsequentialisten in weit höherem Maße, d. h. in einer größeren Anzahl von Entscheidungssituationen, rational, den Forderungen dieser normativen Institution zu genügen. Es ist daher anzunehmen, daß die Zerstörung normativer Institutionen durch konsequentialistisch orientierte Moralität ein schleichender Prozeß wäre.

Es scheint, daß schon den Klassikern des Utilitarismus die langfristige Unvereinbarkeit normativer Institutionen und konsequentialistischer Moralität bewußt war. Die auffallende Tendenz, utilitaristische Handlungskriterien auf die C in foro externo von juridischen Sanktionen etablierten Regeln kollektiver Entscheidungsfindung sowie hierarchischer und demokratischer Kontrolle strukturell geprägten öffentlichen und speziell politischen Entscheidungsbereiche zu beschränken, fände auf diese Weise eine naheliegende Erklärung. Einerseits sind öffentliche Entscheidungen in einem wesentlich geringeren Ausmaß von der Orientierung der Akteure an den im Alltagsleben bewährten normativen Institutionen beeinflußt, so daß ethisch fundierte Handlungskriterien in diesem Bereich in besonderer Weise notwendig erscheinen, andererseits ist die Gefahr einer Zerstörung bewährter normativer Institutionen aufgrund der stärkeren Strukturierung in foro externo geringer.

Die Zerstörung normativer Institutionen durch strikt konsequentialistische moralische Akteure ist vereinbar mit der Kooperativitätseigenschaft des strikten Konsequentialismus bei statischer Betrachtung; denn die Zerstörung normativer Institutionen verändert die Rahmenbedingungen zukünftiger konsequentialistisch rationaler Entscheidungen. Erst unter einem dynamischen Aspekt werden punktuelle und strukturelle Optimierung auch für den strikten Konsequentialismus (nicht nur für abgeschwächte Konsequentialismus-Varianten) unvereinbar[17].

§ 44 *Ziel in intentione recta*

Bisher wurde nur die stärkere Form einer Kritik zweiter Ordnung behandelt: Es ist – konsequentialistisch beurteilt – nicht wünschenswert, daß die Menschen einer Gesellschaft mit bestimmten sinnvollen normativen Institutionen rationale Konsequentialisten werden. Dabei wurde zugunsten

[17] Dieser Befund ist sowohl für die (normative) Kritik erster Ordnung als auch für die Kritik zweiter Ordnung wesentlich. Als Kritik *erster* Ordnung handelt es sich bei dynamischer Betrachtung um die Verletzung der Kooperativitätseigenschaft (s. § 31). Insofern kann dieser Paragraph auch als eine Ergänzung zum Koordinationsproblem des Konsequentialismus gelesen werden. Als Kritik zweiter Ordnung ist die Tatsache, daß eine Gesellschaft rationaler konsequentialistischer Akteure unter konsequentialistischen Kriterien nicht wünschenswert ist, nur dann ein ernsthafter Einwand gegen eine konsequentialistische Ethik, wenn die These vom konstruktivistischen Charakter einer adäquaten normativen Ethik zutrifft. Als Kritik zweiter Ordnung ist das Argument jedoch von normativen Kriterien (wie z. B. der Kooperativitätsbedingung) unabhängig.

des Konsequentialismus angenommen, daß die Maximierung der betreffenden Wertfunktion den einzelnen Personen keine Schwierigkeiten bereite.

Die Wertfunktionen konsequentialistischer Theorien haben einen aggregativen Charakter. Es ist zu ihrer Konstitution unumgänglich, daß von Verfahren der inter- und intrapersonellen Aggregation unterschiedlicher Wertungsaspekte Gebrauch gemacht wird, was voraussetzt, daß inter- und intrapersonelle Bewertungsvergleiche unbegrenzt möglich sind. Die Kritik, die aufgrund dieser starken methodischen Voraussetzungen konsequentialistischer Ethik naheliegt, wurde deshalb nicht erwähnt, weil wir von einem idealtypischen Charakter konsequentialistischer Ethik ausgegangen sind. Diese methodischen Probleme kann man zugunsten des konsequentialistischen Ansatzes als bloße Probleme der Anwendung betrachten. Auch weiterhin sollen diese methodologischen Probleme des Konsequentialismus ausgespart bleiben. Nur ein Aspekt der Konstitutionsproblematik der Wertfunktion ist so gravierend, daß er nicht unerwähnt bleiben kann.

Das Rationalitätsmodell des Konsequentialismus enthält eine subjektive, auf Wertmaximierung ausgerichtete Handlungsorientierung, die sich in einer spezifischen Beziehung motivierender und vorausgehender Absichten niederschlägt, wie sie im ersten Kapitel charakterisiert wurde: Die subjektiven (vorausgehenden) Absichten des idealen Konsequentialisten stehen in einer bestimmten instrumentellen Beziehung zu den motivierenden Absichten, die die (interpersonell invariante) Wertfunktion der betreffenden Theorie des (strikten) ethischen Konsequentialismus konstituieren.

Je nach spezifischer Ausformung des ethischen Konsequentialismus wurde diese Wertfunktion meist als Maß der Verwirklichung von Glück oder Zufriedenheit in der menschlichen Gesellschaft interpretiert. Glück ist jedoch keine geeignete Handlungsmotivation. Menschen, die ihr eigenes Handeln darauf ausrichten, ein möglichst hohes Maß an Zufriedenheit zu erreichen, werden i. allg. wohl ein recht unzufriedenes Leben führen. Ein glückliches Leben ist Nebenfolge der gelungenen Verwirklichung „primärer Ziele"[18]. Glück und Zufriedenheit gehören nicht zu den geeigneten primären Zielen. Dies gilt individuell, aber auch im Rahmen gesellschaftlichen Handelns[19]. Wenn öffentliche Entscheidungen unter dem Aspekt der Glücksmaximierung in intentione recta getroffen werden, werden sie ihr Ziel i. allg. ebenfalls verfehlen. Wenn man diesen anthropologischen Sachverhalt berücksichtigt, muß man wohl vermuten, daß eine Ge-

[18] Vgl. R. Spaemann, „Philosophie als Lehre vom glücklichen Leben", in: *Die Frage nach dem Glück*, hg. von G. Bien, Stuttgart 1978.

[19] Eine sorgfältige Darstellung dieser Problematik bietet O. Höffe, *Strategien der Humanität. Zur Ethik öffentlicher Entscheidungsprozesse*, Freiburg/München 1975, bes. Kap. 7 und 12. Vgl. a. O. Marquard, „Glück im Unglück". Zur Theorie des indirekten Glücks zwischen Theodizee und Geschichtsphilosophie", in: *Die Frage nach dem Glück*, hg. von G. Bien, Stuttgart 1978.

sellschaft rationaler T-Konsequentialisten schon wegen des Fehlens primärer Ziele die Wertfunktion W_T nicht maximiert. Der ethische Konsequentialismus ist also aus elementaren anthropologischen Gründen einer Kritik zweiter Ordnung ausgesetzt, die zwar ebenfalls den konstruktivistischen Charakter des ethischen Konsequentialismus voraussetzt, aber auf eine dynamische Betrachtung verzichten kann.

Teil IV:
Kritik konsequentialistischer Rationalität

14. Kapitel:
Selbstaufhebung konsequentialistischer Rationalität

Der strikte ethische Konsequentialismus ist unterschiedlichen Formen der Kritik ausgesetzt. Ein Typus der Kritik richtet sich gegen die unrealistischen anthropologischen Voraussetzungen, die gemacht werden müssen, um den strikten Konsequentialismus im Sinne einer konstruktivistischen Auffassung auch als praktische Handlungsanleitung zu verstehen. Obwohl dieser Typus der Kritik seine Berechtigung hat, wird er in dieser Untersuchung weitgehend ausgespart, da unser Ziel in erster Linie in der Rekonstruktion und Kritik des *Idealtypus* konsequentialistischer Moral bzw. konsequentialistischer Rationalität liegt. Gerade die einer pragmatischen und anthropologischen Kritik am stärksten ausgesetzte Variante des *strikten* Konsequentialismus ist jedoch – aufgrund ihrer „unrealistischen" Annahmen – als Idealtypus besonders resistent gegenüber einer ihrerseits ebenfalls idealtypischen Kritik. Das letzte Kapitel hat daher die wichtigsten Dimensionen der idealtypischen Kritik des strikten Konsequentialismus entwickelt. Es konnte gezeigt werden, daß wesentliche normative Kriterien – Integrität der Person (Kap. 9), Wahrung individueller Rechte (Kap. 10), interpersonelle Koordination (Kap. 11), Strategie- und Manipulationsunabhängigkeit (Kap. 12), Stabilität normativer Institutionen (§ 27 u. 43) – selbst von dieser theoretisch stärksten Variante des Konsequentialismus nicht erfüllt werden. Auch bei (Selbst-) Anwendung konsequentialistischer Kriterien (Kap. 13) ist der strikte ethische Konsequentialismus, verstanden als konstruktive normative Theorie, unzureichend.

Die Kritik des ethischen Konsequentialismus hat streckenweise die Grenzen zu einer Kritik des Konsequentialismus als Rationalitätsmodell schon im vorangegangenen Kapitel überschritten, besonders deutlich bei der Analyse strategischer Manipulation (Kap. 12). Dies geschah nicht nur aus darstellungstechnischen Gründen. Die Kritik des ethischen Konsequentialismus läßt sich nicht auf eine moralische Dimension beschränken, da für den konsequentialistischen Ansatz ein enger Zusammenhang von Rationalitätsmodell und normativer Ethik charakteristisch ist.

Dieses Kapitel wendet sich daher der Kritik der konsequentialistischen Rationalitätskonzeption selbst zu und kehrt damit an den Anfang der Untersuchung – jetzt allerdings kritisch und nicht rekonstruktiv – zurück. Dabei entfallen die für den ethischen Konsequentialismus typischen Einschränkungen, besonders die der „unparteilichen" – und daher interper-

sonell invarianten – subjektiven Wertfunktionen. Diese Einschränkungen sind nicht nur für den „unrealistischen" Charakter des ethischen Konsequentialismus verantwortlich, sondern sie beheben auch – jedenfalls prima facie – das zentrale Problem konsequentialistischer Rationalität, nämlich das der Koordination individueller Handlungspläne im gesellschaftlichen Kontext.

Koordination hat einen interpersonellen, aber auch einen *intrapersonellen* Aspekt. Selbst wenn sich das interpersonelle Koordinationsproblem durch Sanktionen in foro externo als Ergebnis je individuell maximierenden Verhaltens sicherstellen ließe, bliebe ein strukturgleiches, intrapersonelles Koordinationsproblem. Wir beginnen mit der Analyse des (vertrauteren) interpersonellen Koordinationsproblems, um anschließend zu zeigen, daß es ein strukturgleiches intrapersonelles Koordinationsproblem gibt. Koordination ist aber nicht der einzige Aspekt der Selbstaufhebung des Konsequentialismus. Das angesprochene Problem des Handelns im Hinblick auf Ziele in intentione recta ist ein weiterer Bestandteil der Selbstaufhebung des Konsequentialismus als Rationalitätsmodell. Allerdings steht es in einem engen Zusammenhang mit dem Koordinationsproblem.

§ 45 Kollektive Irrationalität

Es gibt traditionsgemäß zwei Gegenstände der ethischen Debatte:

Erstens die Frage danach, was (intrinsisch) wertvoll ist bzw. welchen Zielen menschliches Handeln dienen sollte. Mit anderen Worten: Welches Leben ist gut?

Zweitens, angenommen, wir wissen, was gut ist, welches Handeln ist dann (moralisch) richtig?

Sowohl der klassischen Ethiktradition, wie sie von Platon und Aristoteles begründet wurde, als auch der konsequentialistischen Theorie der Rationalität ist die Vorstellung gemeinsam, die zweite Frage sei mit der Beantwortung der ersten als erledigt zu betrachten. Für die konsequentialistische Theorie der Rationalität geschieht dies in Form des Erwartungswertmaximierungskriteriums. Die hier vertretene Gegenthese lautet, daß die Beantwortung der zweiten Frage weitgehend unabhängig von der Beantwortung der ersten Frage ist; denn das Koordinationsproblem erfordert eigenständige Handlungskriterien, die sich nicht als bloße Maximierungskriterien derjenigen subjektiven Wertungen rekonstruieren lassen, die die jeweilige Entscheidungssituation charakterisieren.

In manchen Entscheidungssituationen führt konsequentialistisch rationales Verhalten der einzelnen Akteure zu einem Ergebnis, das den subjektiven Wertungen bzw. Präferenzen jedes einzelnen beteiligten Akteurs weniger entspricht als ein Ergebnis, das zu erreichen gewesen wäre, wenn

sich die Beteiligten anders, aber nicht konsequentialistisch rational entschieden hätten. Besonders interessant sind darunter diejenigen Situationen, in denen sich dieser Defekt nicht durch zusätzliche Information beheben läßt. Ein prominentes Beispiel einer solchen Situation ist das sogenannte *Gefangenendilemma.* Es trägt diesen Namen aufgrund einer Geschichte, die wir hier kurz wiederholen: Zwei Personen werden, nachdem sie einen bewaffneten Raubüberfall begangen haben, festgenommen und in getrennten Zellen verhört. Wenn beide ihre Tat ableugnen, können sie nicht überführt werden und werden wegen unerlaubten Besitzes und Tragens von Waffen zu einem Jahr Gefängnis verurteilt. Wenn beide ihre Tat gestehen, werden sie wegen bewaffneten Raubüberfalls zu je fünf Jahren verurteilt. Wenn jedoch einer der beiden die Tat gesteht und der andere nicht, dann kommt der Geständige als Kronzeuge der Anklage ohne Bestrafung frei, während der Nichtgeständige zu sechs Jahren Gefängnis verurteilt wird.

Anhand dieser einfachen und viel diskutierten Interaktionssituation läßt sich das Problem der Kooperation rationaler Konsequentialisten deutlich machen. Es gibt vier mögliche Konsequenzen der Interaktionssituation, die durch eine bestimmte Verteilung von Haftjahren auf die beiden Gefangenen charakterisiert sind. Wenn wir annehmen, daß jeder der beiden Gefangenen diese Konsequenzen ausschließlich danach beurteilt, wieviel Zeit er im Gefängnis verbringen muß, dann hat der eine der beiden Gefangenen (A) folgende Rangordnung für die vier möglichen Ergebnisse des Verhörs: Am liebsten ist ihm, daß B nicht gesteht und er – als Kronzeuge der Anklage – durch ein Geständnis straffrei ausgeht; an zweiter Stelle der subjektiven Rangordnung steht, daß beide nicht gestehen, und A ein Jahr Gefängnisstrafe erhält; an dritter Stelle folgt die Option, daß beide gestehen und A fünf Jahre in Kauf nehmen muß; und zuletzt die für A ungünstigste Möglichkeit, daß B gesteht und als Kronzeuge der Anklage frei kommt, während A sechs Jahre abbüßt.

Als konsequentialistisch rationale Person hat A daher folgende Handlungspräferenzen: Wenn A annehmen kann, daß B nicht gesteht, dann wird A dennoch gestehen, um als Kronzeuge der Anklage straffrei auszugehen (würde A nicht gestehen, hätte er immerhin ein Jahr Gefängnis abzubüßen). Wenn A annehmen muß, daß B gesteht, dann wird A ebenfalls gestehen, um seine Gefängnisstrafe von sechs auf fünf Jahre zu vermindern. Im Sinne des konsequentialistischen Entscheidungskriteriums ist die Handlungsalternative „Gestehen" für A in dieser Situation *dominant:* D. h., unabhängig von den eintretenden Umständen ist es für A konsequentialistisch rational zu gestehen.

Gleiches gilt aufgrund der Symmetrie der Interaktionssituation auch für B. Unabhängig davon, wie sich der andere jeweils entscheidet, ist es für jeden der beiden Gefangenen konsequentialistisch rational, die Tat zu gestehen. Wenn beide Gefangenen konsequentialistisch rational handeln, wer-

den beide gestehen und fünf Jahre ins Gefängnis kommen, obwohl sie doch – hätten beide geschwiegen – auch mit einem Jahr Gefängnisstrafe hätten davonkommen können.

Man kann dieses Ergebnis aus moralischen Gründen befürworten und selbst die Institution des Kronzeugen gutheißen, da sie konsequentialistisch rationalen Akteuren Geständnisse abringt; für die konsequentialistische Rationalitätskonzeption ist dieses Ergebnis jedoch unbefriedigend: Immerhin stand beiden Akteuren jeweils eine Handlung offen, die beide besser gestellt hätte, wenn diese Alternative übereinstimmend gewählt worden wäre. Ein wohlwollender Berater der beiden hätte ihnen folglich eingeschärft, kein Geständnis zu machen. Wäre den Gefangenen angeboten worden, daß derjenige mit einer Sanktion belegt wird, der durch sein Geständnis die günstige Möglichkeit einer kurzen Haftzeit vereitelt, dann hätten beide diesem Angebot zugestimmt. Wenn nur zwei Kombinationen individueller Handlungen – beide Gefangenen gestehen oder beide Gefangenen gestehen nicht – in Betracht kämen, dann wäre „nicht gestehen", auch bei konsequentialistischer Betrachtung, je indidivuell besser. Für jeden einzelnen ist es jedoch konsequentialistisch rational, zu gestehen, da ihm zusätzlich die asymmetrischen Optionen offenstehen. Die Symmetrie der Situation verlangt andererseits, daß sich beide Personen, sofern sie gleichermaßen rational sind, symmetrisch verhalten.[1]

Als normative Theorie *empfiehlt* die konsequentialistische Rationalitätskonzeption je individuell Entscheidungen, die – von mehreren Personen ausgeführt – in einigen Entscheidungssituationen Konsequenzen haben, die *je individuell* nachteilig sind. Wenn die Beteiligten ideale rationale Personen im Sinne des Konsequentialismus sind, dann wünschen sie sich zwar, daß jede einzelne Person gezwungen wäre, sich kooperativ zu verhalten, und dennoch verhält sich jede von ihnen ohne einen solchen Zwang unkooperativ. Ein jeder wünscht sich, daß sich alle kooperativ verhalten, aber jeder einzelne verhält sich unkooperativ. Als eine Theorie idealer Entscheidungsfindung im sozialen Kontext ist das konsequentialistische Rationalitätsmodell daher wenig attraktiv. Präziser formuliert: Als *individuelle* Handlungsanweisung verstanden, ist das konsequentialistische Kriterium in einer Situation von der Art des Gefangenendilemmas *auch nach konsequentialistischem Maßstab* erfolgreich: Wer sich daran hält, maximiert je individuell seine subjektive Wertfunktion; d. h., er hätte je *individuell* nichts anderes tun können, das ein besseres Ergebnis gezeigt hätte. In einer Situation von der Art des Gefangenendilemmas hätte jedoch ein anderes, nicht-konsequentialistisches Entscheidungskriterium – *allgemein befolgt* – für *jedes* Mitglied dieser Gruppe bessere Konsequenzen gehabt,

[1] Diese Problematik wird von mir detaillierter behandelt in „Das rational choice-Paradigma – Extensionen und Revisionen", in: *Praktische Rationalität. Grundlagenprobleme und ethische Anwendungen des rational choise-Paradigmas*, hg. von J. Nida-Rümelin u. U. Wessels, Berlin/New York 1993.

als die allgemeine Befolgung des konsequentialistischen Entscheidungskriteriums. Wenn eine Gruppe von konsequentialistisch rationalen Personen vor der Wahl stünde, unter einer Reihe von Entscheidungskriterien das für sie günstigste auszuwählen, dann würde das konsequentialistische Kriterium nicht gewählt werden. Diese Annahme hat als einzige Prämisse die Pareto-Inklusivität der kollektiven Rangordnung[2] – eine Prämisse, die von allen gängigen Wahlregeln und natürlich auch von der Mehrheitswahlregel erfüllt wird. Dieser Sachverhalt rechtfertigt die Feststellung: *Die generelle Befolgung konsequentialistischer Entscheidungskriterien ist im allgemeinen kollektiv irrational*[3]. Konsequentialistische Theorien, die Interaktionssituationen von der Art des Gefangenendilemmas *nicht ausschließen*, sind daher als konstruktive Theorien i. allg. inadäquat. Konsequentialistische Theorien, die Interaktionssituationen von der Art des Gefangenendilemmas *ausschließen*, sind dagegen aus anderen Gründen inadäquat. Einige dieser Gründe wurden in Teil III dargelegt. Die prominenteste konsequentialistische Theorie, die Interaktionssituationen von der Art des Gefangenendilemmas ausschließt – der strikte ethische Konsequentialismus – wurde dabei besonders eingehend analysiert.

Es gibt für konsequentialistische Theorien unterschiedliche Möglichkeiten, Interaktionssituationen von der Art des Gefangenendilemmas zu umgehen. Interpersonell invariante Wertfunktionen stellen nur eine (recht radikale) Möglichkeit dar. Weitere Möglichkeiten bietet die Etablierung von Sanktionen (wir nennen sie die Hobbes-Variante[4]), die Annahme altruistischer Handlungsmotive (Shaftesbury-Variante[5]) sowie die anthropologische Prämisse einer dispositionellen Vorprägung, die auch das Verhalten konsequentialistisch rationaler Personen bestimmt (Mackie-Variante[6]).

Nehmen wir an, die Gefangenen entscheiden sich beide, ohne Absprache, nicht zu gestehen. Wir können auch sagen, sie entscheiden sich beide für die kooperative Strategie. Für den Konsequentialisten wäre dies deshalb irrational, weil beide Personen – jedenfalls im Sinne einer Minimierung ihrer jeweiligen Gefängniszeit – nicht erwartungswertmaximierend gehandelt haben. Wenn wir die Wahl der kooperativen Strategie „nicht gestehen" deshalb als erklärungsbedürftig betrachten, dann lassen sich die folgenden fünf Typen von Erklärungen unterscheiden:

[2] Zur Definition der Pareto-Inklusivität s. § 29.

[3] *„Generell"* deshalb, weil die Befolgung des konsequentialistischen Entscheidungskriteriums einer Theorie T nicht in allen, sondern nur in einigen Interaktionssituationen suboptimal im Sinne der Wertfunktion von T ist; *„im allgemeinen"*, weil diese Feststellung nur für konsequentialistische Theorien zutrifft, die Interaktionssituationen von der Art des Gefangenendilemmas zulassen. Der strikte ethische Konsequentialismus und alle konsequentialistischen Theorien mit interpersonell invarianter Wertfunktion (also z. B. auch Theorien des Gerechtigkeitskonsequentialismus) schließen Situationen dieser Art aus, vgl. Kap. 5–7.

[4] Th. Hobbes, *De cive* (1647), (engl. 1651), 2. Teil.

[5] Shaftesbury, *An Inquiry concerning Virtue or Merit*, London, 1711.

[6] J. L. Mackie, *Ethics*, Harmondsworth 1977, Kap. 8.

Typ 1 (Irrationalität)

Das (kooperative) Verhalten wird mit *Informationsdefiziten* bzw. *epistemischer* oder *präferenzieller Irrationalität* erklärt, was z. B. durch die Unkenntnis der tatsächlichen Folgen einzelner Handlungsalternativen, durch die geringe Intelligenz der Beteiligten, die es ihnen verwehrt, die Struktur der Entscheidungssituation zu durchschauen, durch die Spontaneität ihrer Entscheidung oder die Suggestionskraft des ermittelnden Beamten verursacht sein könnte.

Typ 2 (Shaftesbury-Variante)

Das (kooperative) Entscheidungsverhalten wird damit erklärt, daß neben dem egoistischen Motiv, möglichst wenig Gefängnisjahre absitzen zu müssen, auch *altruistische Motive*, etwa das, der anderen Person keinen Schaden zuzufügen, oder Mitleid eine Rolle spielen.

Typ 3 (Hobbes-Variante)

Man zieht zur Erklärung des (kooperativen) Verhaltens den Umstand heran, daß den Kronzeugen der Anklage außerhalb des Gefängnisses gewisse *Sanktionen* erwarten, wie sie z. B. für das organisierte Verbrechen typisch sind.

Typ 4 (Mackie-Variante)

Man erklärt das (kooperative) Verhalten mit der starken *Disposition* der beiden Gefangenen, die Wahrheit zu sagen, und macht die Wahl der richtigen Disposition zum Gegenstand der rationalen Entscheidung.

Typ 5 (Kant-Variante)

Die Personen orientieren sich an *kollektiver Rationalität* – bzw. Kantisch formuliert: Sie achten darauf, daß die Maxime ihres Handelns jederzeit (in dieser kleinen Welt) zur Grundlage einer allgemeinen Gesetzgebung werden könnte; das gilt aber nur für das kooperative, nicht für das unkooperative Verhalten in dieser Entscheidungssituation.

Die erste und die letzte Erklärungsvariante sind nicht-konsequentialistisch, während die übrigen Varianten geeignet sind, die kooperative Strategie mit einer konsequentialistischen Rationalitätskonzeption zu versöhnen.

Der erste Typ ist für meine Fragestellung irrelevant. Natürlich gibt es zahlreiche Formen von Irrationalität, und es mag sogar sein, daß die Mehrzahl der Entscheidungen des Alltags auch im Sinne einer idealen Theorie der Rationalität als irrational zu charakterisieren sind, aber für eine Kritik des konsequentialistischen Rationalitätsmodells würde es ausreichen, eine einzige Situation von der Art des Gefangenendilemmas zu finden, bei der rationale Personen sich kooperativ verhalten[7].

[7] Das Problem der Kooperation ist ein zentrales Thema der Gesellschaftswissenschaften. In

Typ 1 und 5 haben etwas gemeinsam: Es wird nicht angezweifelt, daß die tatsächliche Entscheidungssituation die Struktur eines Gefangenendilemmas hat, während die Erklärung der übrigen Typen gerade darin besteht, aufzuzeigen, daß es sich nur scheinbar um eine Gefangenendilemma-Situation handelt.

Ohne Zweifel sind diese Erklärungsvarianten 2, 3 und 4 in vielen Fällen angemessen: Altruistische Motive können eine Entscheidung konsequentialistisch rational machen, die bzgl. der eigenen Interessen nicht rational wäre (Typ 2); zu erwartende Sanktionen können eine Entscheidung konsequentialistisch rational machen, die ohne diese Sanktionen nicht rational wäre (Typ 3); starke dispositionelle Prägungen können die Entscheidungsfreiheit einer Person dergestalt einschränken, daß prima facie offenstehende Handlungen ausgeschlossen sind. In allen drei Fällen würde eine vollständige Analyse der Entscheidungssituation ergeben, daß es sich in Wirklichkeit nicht um eine Gefangenendilemma-Situation handelt, denn „nicht gestehen" maximiert, wenn eine dieser Varianten zutrifft, auch individuell die jeweilige Wertfunktion. Bei Typ 2 ergibt sich dies daraus, daß die Wertfunktion nicht monoton mit je individuell abnehmender Gefängniszeit wächst, sondern eine Funktion der Verteilung der Gefängnisjahre auf beide Personen wird. Bei Typ 3 geben die Gefängnisjahre die entscheidungsrelevanten individuellen Folgen nur unvollständig wieder. Der Vorteil, die eigene Gefängniszeit zu verringern, wird durch persönliche Nachteile anderer Art je individuell mehr als aufgewogen. Bei Typ 4 gehört die Alternative „gestehen" nicht zu den offenstehenden Strategien[8].

Die für die konsequentialistische Rationalitätskonzeption entscheidende Frage lautet jedoch: „Kann es eine Situation geben, in der die folgenden Aussagen zugleich zutreffen?"

1) Die Gefangenen gestehen nicht.

2) Es ist das einzige Ziel der Gefangenen, die Anzahl ihrer jeweiligen Gefängnisjahre zu minimieren.

3) Die Gefangenen entscheiden rational.

Fortsetzung Fußnote von Seite 156

der Soziologie und politischen Theorie dominiert der Erklärungstyp 3 (Hobbes). Der ökonomischen Theorie der Kooperation kommen Gefangenendilemma-Situationen ungelegen, so daß Kooperation meist dadurch erklärt wird, daß aufgrund von Iteration in Wirklichkeit keine Gefangenendilemma-Situation vorliege (vgl. z. B. Axelrod 1988) oder eine Fehlinterpretation der subjektiven Präferenzen ausschlaggebend sei (Shaftesbury). In der Sozialpsychologie dominieren Beiträge vom Erklärungstyp 1 und in der Ethik Beiträge vom Erklärungstyp 4 und 5.

[8] Dies mag im Einzelfall die richtige Analyse sein, als generelle Lösung des Kooperationsproblems ist diese Variante aber nicht geeignet. Vgl. meine Kritik an der Konzeption Gauthiers, der seine ethische Theorie ebenso wie J. L. Mackie auf eine dispositionelle Lösung des Kooperationsproblems stützt: D. Gauthier, *Morals by Agreement*, Oxford 1986 und Verf., „Practical Reason, Collective Rationality and Contractarianism", in: *The Contractarian Enterprise*, hg. von R. Sudgen, London 1990, i. E.

4) Die Gefangenen könnten jederzeit auch lügen oder ihr Geständnis zurückhalten. Sie sind nicht (dispositionell) gezwungen, ein Geständnis zu machen.

5) Außerhalb des Gefängnisses erwarten keinen der beiden Gefangenen weitere Sanktionen.

Wenn diese Frage mit „Ja" beantwortet wird, dann ergeben sich folgende Implikationen: Irrationalität ist durch (3) ausgeschlossen. (2) schließt die Shaftesbury-Variante ebenso wie eine Orientierung an einem strikten ethischen Konsequentialismus aus, (4) die Mackie-Variante und (5) die Hobbes-Variante. Es bleibt nur die nicht-konsequentialistische Erklärungsvariante[9]. Niemand wird bestreiten, daß die Aussagen (1), (2), (4) und (5) zugleich wahr sein können. Intuitiv spricht nichts dagegen, daß daneben auch (3) zutreffen kann. Dies wäre allerdings mit einer konsequentialistischen Rationalitätskonzeption nicht mehr vereinbar. Die Beweislast liegt zunächst bei demjenigen, der eine intuitiv überzeugende These angreift – in diesem Fall also bei dem Verfechter einer konsequentialistischen Rationalitätskonzeption. Dennoch gewinnt die Kritik an Überzeugungskraft, wenn sich dieses Gegenbeispiel als Spezialfall eines tiefer liegenden Defizits jeder konsequentialistischen Rationalitätskonzeption aufweisen läßt. Dies ist der Gegenstand der folgenden Abschnitte.

§ 46 Rationalität und praktische Vernunft

In der Literatur ist eine Rationalitätskonzeption weit verbreitet, die man etwa folgendermaßen charakterisieren könnte: Eine Entscheidung (oder in unserer Terminologie: eine vorausgehende Absicht) ist rational, wenn sie den Zielen des Handelnden in optimaler Weise entspricht. Wir nennen diese Rationalitätskonzeption *„subjekt-zentriert"*. Die Relativierung auf die Ziele des Handelnden schließt für die meisten objektivistischen ethischen Theorien eine normative Interpretation aus, d. h., ein Rationalitätskriterium dieser Art kann nicht durch weitere einschränkende Bedingungen zum normativen Kriterium einer objektivistischen ethischen Theorie werden. Für eine Ethik von Kantischem Typ würde ein Rationalitätskriterium dieser Art als normatives Kriterium schon deshalb nicht in Frage kommen, weil es einen *hypothetischen* Charakter hat. Für subjektivistische Ethiken sind kategorische Kriterien jedoch ohnehin ausgeschlossen – eine normative Interpretation im subjektivistischen Rahmen ist daher nicht nur möglich, sondern auch naheliegend. Tatsächlich spielt für subjektivistische Ethiken die Rationalität der Befolgung von Normen bzw. der Nachweis, daß es im Interesse des Handelnden ist, sich an bestimmte moralische Regeln zu halten, in doppelter Weise eine große Rolle: einmal als

[9] Vgl. Verf.: „Practical Reason or Metapreferences? An Undogmatic Defense of Kantian Morality", *Theory and Decision* 30 (1991) 133–162.

eine anthropologische Prämisse, daß Personen ihr Handeln, i. allg. an ihren persönlichen Interessen ausrichten, und zum anderen als Begründungsmethode: Eine moralische Regel ist erst dann ethisch begründet, wenn ihre Etablierung und Befolgung im Interesse aller oder der in Frage stehenden Gruppe von Personen ist[10].

(Praktische) Vernunft darf nicht mit einer so verstandenen Entscheidungsrationalität in eins gesetzt werden: Eine Handlung bzw. Entscheidung ist vernünftig, wenn sie gegenüber jedermann gerechtfertigt werden kann, aber nicht jede Entscheidung, die den Zielen des Handelns optimal entspricht, kann gegenüber jedermann gerechtfertigt werden. Das gilt besonders dann, wenn diese Ziele mit individuellen Rechten anderer Personen kollidieren.

Betrachten wir folgendes Beispiel:

Onkel Karl, ein älterer, reicher Junggeselle, und Fritz, sein Neffe und einziger Erbe, sitzen in einem kleinen Boot und angeln. Das Gewässer wimmelt von Haifischen. Kein anderes Boot ist in Sicht. Es entspinnt sich folgender Dialog:

Fritz: Weißt du, ich will um alles in der Welt zu einer halben Million Dollar kommen, um sie für meine Vergnügungen auszugeben.

Onkel Karl: Ich kenne dich gut genug, um dir das zu glauben. Du hast in letzter Zeit ja eine Reihe von Versuchen gemacht, um an Geld zu kommen, die aber wohl alle schiefgegangen sind.

Fritz: So ist es. Ich bin unterdessen zu der Überzeugung gekommen, daß es für mich nur eine Möglichkeit gibt, zu Geld zu kommen – ich bringe dich um.

Onkel Karl: Ich fürchte, du hast bei deinen begrenzten Begabungen und deinem Hang zur Faulheit tatsächlich keine andere Möglichkeit, um zu Geld zu kommen.

Fritz: Da du zugibst, daß es für mich nur eine Möglichkeit gibt, zu Geld zu kommen: Welchen Rat gibst du mir?

Onkel Karl: Da du nur einen Wunsch hast, nämlich rasch zu Geld zu kommen, und es für dich keine andere Möglichkeit gibt, als mich umzubringen, solltest du mich aus dem Boot stoßen.

Fritz: Ich stimme beiden Prämissen und deiner Schlußfolgerung zu. Da ich den guten Rat meines Onkels immer befolge, werde ich dich jetzt aus dem Boot stoßen. (Fritz stößt Onkel Karl aus dem Boot.)[11]

[10] Vgl. z. B. N. Hoerster, „Moralbegründung ohne Metapyhsik", *Erkenntnis* 19 (1983) 225–238.

[11] Dieses Beispiel ist sinngemäß einem Artikel von R. M. Hare entnommen, wo es allerdings in einem anderen Kontext steht. Hare argumentiert damit gegen Black für das sog. Humesche Gesetz (ein Soll-Satz läßt sich nicht aus Ist-Sätzen allein ableiten). Vgl. M. Black, „The Gap between ‚Is' and ‚Should'", *Philosophical Review* 73 (1964) 165–181; R. M. Hare,

Das Beispiel ist nicht damit abgetan, daß Fritz' Handlung, den Onkel vom Boot zu stoßen, unmoralisch ist und deshalb von Onkel Karl auch nicht empfohlen werden dürfe. Denn in vielen Fällen versteht man Ratschläge und Empfehlungen nicht als Gebote der Moral, sondern als solche der Klugheit (oder Rationalität). Dieser Dialog läßt jedenfalls keine andere Interpretation zu: Fritz möchte nicht wissen, was *moralisch geboten* ist, sondern was er – in wohlverstandenem Eigeninteresse – tun soll. Nun könnte man meinen, der Dialog kranke ausschließlich daran, daß er psychologisch wenig plausibel ist, da es völlig unrealistisch sei, daß Onkel Fritz empfiehlt, ihn zu töten. Aber auch darin liegt nicht die philosophisch interessante Problematik dieses Beispiels.

Es kann sein, daß sowohl die intentionale (konative, präferentielle) Prämisse: Fritz hat den dominanten Wunsch, zu Geld zu kommen, als auch die epistemische Prämisse: Fritz weiß, daß es für ihn keine andere Möglichkeit gibt, an Geld zu kommen, wahr ist, so daß es unbestreitbar zutrifft, daß es in Fritz' Interesse liegt, den Onkel vom Boot zu stoßen, ohne daß die Conclusio „Fritz sollte Karl vom Boot stoßen" gerechtfertigt ist. Die beiden Prämissen scheinen für diese Conclusio nicht hinzureichen. Natürlich ist es möglich, daß ein Freund, der die Geisteshaltung von Fritz teilt, in Anbetracht der von den beiden Prämissen formulierten Sachverhalte Fritz empfiehlt, er solle den Onkel töten. Aber damit geschieht etwas Bemerkenswertes: Mit dieser Empfehlung identifiziert sich der Freund in einem gewissen Sinne mit den Zielen von Fritz.

Onkel Karl unterliefe jedenfalls kein logischer Fehler, wenn er beide Prämissen als empirisch zutreffend anerkennt, ohne die Conclusio zu akzeptieren. Onkel Karl könnte, um den Fallstricken sprachlicher Unschärfe zu entgehen, antworten: „Ich bezweifle nicht, daß beide Annahmen zutreffen. Daher weiß ich, daß du dein Ziel nur erreichen kannst, wenn du mich vom Boot stößt. Aber ich meine nicht, daß du deine Ziele unter diesen Umständen verfolgen solltest. Deshalb bin ich auch nicht der Auffassung, daß du mich aus dem Boot stoßen solltest." Es ist nicht erforderlich, daß die Rede „du solltest" in dieser Äußerung moralisch interpretiert wird. Wenn Onkel Karl trotz der für ihn negativen Folgen der Auffassung ist, Fritz sollte seine Ziele auch unter diesen Umständen verwirklichen, dann könnte er der Conclusio zustimmen, ohne damit einen moralischen Anspruch zu formulieren.

Es gibt gute Argumente gegen die These, daß es eine logische Beziehung zwischen Wünschen, Annahmen und Handlungen – bzw. zwischen motivierenden und vorausgehenden Absichten – gebe. Ganz analog gibt es keine logische Beziehung zwischen abgetrennten „Soll-Sätzen" und epistemischen sowie intentionalen Prämissen. Die Qualifizierung „abgetrennte

Fortsetzung Fußnote von Seite 159

„Wollen: Einige Fallen", in: *Analytische Handlungsphilosophie* Bd. 1, hg. von G. Meggle, Frankfurt/M. 1985.

Soll-Sätze" ist notwendig, weil es einen Gebrauch von „sollen" gibt, der nicht im Sinne einer Handlungsanweisung zu verstehen ist. „Sollen" bedeutet dann nichts anderes als in den Prämissen zum Ausdruck kommt. Ein Soll-Satz dieser Art ist nicht abtrennbar und hat damit einen rein deskriptiven Charakter: Wenn Fritz zu Geld kommen will, und es keine andere Möglichkeit gibt, an Geld zu kommen, als Onkel Karl vom Boot zu stoßen, dann mag man unter Umständen auch davon sprechen, daß Fritz, wenn er zu Geld kommen will, Onkel Karl vom Boot stoßen sollte. Aber dieses „sollte" bedeutet nichts anderes, als daß es für Fritz *keine andere Möglichkeit gibt, sein Ziel zu erreichen*. Niemand wird behaupten, Fritz sollte, angesichts der obwaltenden Umstände, Onkel Karl vom Boot stoßen.

Daraus läßt sich folgern, *daß subjekt-zentrierte Rationalitätskriterien nicht normativ verstanden werden dürfen*. Das subjekt-zentrierte Rationalitätskriterium bleibt im Sinne der Nichtabtrennbarkeit des Sollens-Satzes hypothetisch. Ein Kriterium dieser Art ist trivial. Die Verwendung des Verbums „sollen" verführt dazu, die tiefe Kluft zwischen einer normativen (und damit abtrennbaren) Verwendung und einer deskriptiven Verwendung zu verschleiern. In deskriptiver Verwendung bedeutet „Fritz sollte Onkel Karl aus dem Boot stoßen" nichts anderes als „Fritz hat keine andere Möglichkeit, sein Ziel zu erreichen". Natürlich kann der Sachverhalt, daß Fritz keine andere Möglichkeit hat, sein Ziel zu erreichen, das normative Urteil, daß Fritz den Onkel aus dem Boot stoßen sollte, *rechtfertigen*: z. B. wenn es das Ziel von Fritz ist, das Leben von Schiffbrüchigen zu retten, die sonst ertrinken würden. Wenn das normative Urteil moralisch aufzufassen ist, scheint es eine einfache Folgerung zu geben: Nur wenn der Sprecher die Ziele von Fritz unter einem moralischen Aspekt für gut (oder akzeptabel) hält, können die empirischen Prämissen das normative Urteil begründen. Daraus darf jedoch nicht gefolgert werden, daß sich dann das normative Urteil logisch aus diesen Prämissen ableiten läßt. Die Frage ist, ob jemand darüber hinaus auch in einem außermoralischen Sinne das normative Urteil, daß Fritz Onkel Karl aus dem Boot stoßen sollte, mit dem Umstand rechtfertigen kann, daß Fritz keine andere Möglichkeit habe, sein Ziel zu erreichen. Offenbar würde dies voraussetzen, daß die so urteilende Person die Ziele von Fritz teilt oder genauer: befürwortet, daß Fritz seine Ziele erreicht (was nicht unbedingt voraussetzt, daß die Ziele von Fritz auch seine eigenen sind).

Kann die urteilende Person davon überzeugt sein, daß nur ihr persönlich die Ziele von Fritz akzeptabel erscheinen? Sicherlich kann eine dritte Person ihr normatives Urteil als Ausdruck einer gemeinsamen subjektiven Wertorientierung ansehen und zugleich überzeugt sein, daß sich diese Wertorientierung moralisch nicht aufrechterhalten läßt. Die Person, die urteilt, Fritz solle Karl vom Boot stoßen, muß eine Antwort auf die Frage geben können, warum ihrer Auffassung nach Fritz seine Ziele erreichen

sollte. Wenn die Antwort z. B. lautet: „Weil ich sie zufällig teile", dann verwandelt sich das vermeintlich normative Urteil unter der Hand wieder in ein deskriptives: Die Person hat dann mitgeteilt, daß es für Fritz keine andere Möglichkeit gibt, sein Ziel zu erreichen, als Onkel Karl vom Boot zu stoßen – *und* daß sie darüber hinaus wünscht, daß Fritz sein Ziel erreicht. Die Abtrennbarkeit des normativen Urteils von den deskriptiven Prämissen, die Fritz betreffen, bleibt zwar bestehen, der hypothetische Charakter des normativen Urteils jedoch überträgt sich dann auf die deskriptiven Prämissen, die die urteilende Person betreffen, ohne daß die Möglichkeit der Abtrennung gegeben ist.

Wir haben oben dem Begriff der subjekt-zentrierten Rationalität den der *praktischen Vernunft* gegenübergestellt. Dies könnte nach dem Gesagten als die Gegenüberstellung einer deskriptiven und einer normativen Betrachtungsweise verstanden werden: Fragen der Rationalität als bloße Fragen der Klugheit und Fragen der Vernunft als Fragen der Moral[12]. Die Entscheidungstheorie könnte dann als ein Versuch der Explikation des Rationalitätsbegriffs und die Ethik als Theorie oder Teiltheorie praktischer Vernunft verstanden werden.

Diese Gegenüberstellung hat eine zentrale Schwäche: Sie ist mit der Vorstellung von der *Einheitlichkeit* einer Person nur schwer vereinbar. Wer eine bestimmte Handlung vollzieht und meint, daß sie rational sei, der wird diese Handlung auch gegenüber jedermann zu verteidigen suchen. Mit anderen Worten: Er beansprucht, diese Handlung sei vernünftig. Vice versa: Wer einen bestimmten Typus von Verhalten für gesellschaftlich untragbar hält, ihn jedoch zugleich aus Gründen der Rationalität praktiziert, der erscheint uns als eine „gespaltene" Persönlichkeit. Vollends problematisch werden diese Unterscheidungen angesichts des Aspektes der *Handlungsgründe*. Entweder ist die Bezugnahme auf die eigenen Handlungsziele ein hinreichender Grund, eine bestimmte Handlung zu befürworten (bzw. eine entsprechende vorausgehende Absicht auszubilden), dann kann es darüber hinaus keine Vernunftgründe mehr geben, oder die Rationalität einer Entscheidung ist nur eine von mehreren möglichen Handlungsgründen, so daß es gewichtigere Handlungsgründe daneben geben kann, die dazu raten, die allein rationale Entscheidung nicht zu treffen. In diesem Fall würden Vernunftgründe dafür sprechen, eine *irrationale* Entscheidung zu treffen.

Obwohl nach wie vor solche begrifflichen Unterscheidungen zwischen Rationalität und Vernunft weit verbreitet sind (in den meisten Fällen natürlich nur implizit), drängen die angesprochenen Paradoxien dieser Gegenüberstellung zu einer Auflösung, welche typischerweise zwei entgegen-

[12] Vgl. z. B. D. Richards, *Theory of Reasons for Action*, Oxford 1971, S. 76: „Questions of rationality involve the agents aims and the best way to realize them, whereas questions of reasonableness involve the accessment the persuit of one's own aims in the light of the morality justified claims of the others."

gesetzte Formen annimmt: Die eine Auflösung besteht darin, den Vernunftbegriff im Rationalitätsbegriff aufgehen zu lassen und nur das für wohlbegründet zu halten, was angesichts der Ziele des Handelnden optimal ist. Dieser Weg wird insbesondere von subjektivistischen Ethikern eingeschlagen[13]. Die andere Auflösung besteht in der Annäherung des Rationalitätsbegriffes an den der Vernunft. Bereits die Ausweitung der Entscheidungstheorie auf die Spieltheorie enthält implizit ein Publizitätskriterium, das die Rationalitätskriterien der Spieltheorie gegenüber allgemeiner Offenlegung immun macht. Dieses Publizitätskriterium ist aber zugleich ein wesentliches Charakteristikum praktischer Vernunft.

Offensichtlich gibt es nur zwei Möglichkeiten:

Rationalitätskriterien können im Sinne eines nicht-abtrennbaren Sollens-Satzes (trivial) deskriptiv verstanden werden, oder sie sind als normative Kriterien den Rechtfertigungsansprüchen praktischer Vernunft zu unterwerfen.

Nur die zweite Möglichkeit wird dem konsequentialistischen Rationalitätsverständnis gerecht. Damit aber gewinnt die Tatsache, daß konsequentialistische Handlungsnormen in einigen (vor allem moralisch bedeutsamen) Entscheidungssituationen kollektiv irrationale Ergebnisse zeitigen, einen besonderen Stellenwert. Da gilt, daß in manchen Entscheidungssituationen konsequentialistisch rationales Verhalten der einzelnen Akteure zu einem Ergebnis führt, das den subjektiven Wertungen bzw. Präferenzen jedes einzelnen beteiligten Akteurs weniger entspricht als ein anderes Ergebnis, das zu erreichen gewesen wäre, wenn sich die Beteiligten anders – aber nicht konsequentialistisch rational – entschieden hätten, läßt sich die Befolgung der konsequentialistischen Handlungsregel durch die beiden Gefangenen *kollektiv nicht rechtfertigen*. Dies kann deshalb nicht bestritten werden, weil beide Gefangenen konsequentialistisch günstigere Handlungsfolgen hätten herbeiführen können, wenn sie sich nicht an einem konsequentialistischen Handlungskriterium orientiert hätten.

Die Frage, die sich jetzt darüber hinaus stellt, lautet: Kann die individuelle Wahl der nicht-kooperativen Entscheidung gegenüber jedermann gerechtfertigt werden? Handelt es sich hierbei um eine vernünftige Entscheidung? Prüfen wir, inwiefern sie sich gegenüber dem anderen Gefangenen B im Gefangenendilemma rechtfertigen läßt: B ist offensichtlich in einer analogen Situation wie Onkel Karl. Die Rechtfertigung der nicht-kooperativen Entscheidung wäre nur dann gelungen, wenn B als (epistemisch) rationale Person zugeben muß, daß A richtig handelt, wenn er gesteht. B müßte (als rationale Person) zu A sagen können: „Ja, wenn es dein einziges Ziel ist, deine Gefängnisjahre zu minimieren, dann solltest du gestehen", wobei dieses „Sollen" abtrennbar sein muß. Die Abtrennbarkeit

[13] Vgl. J. L. Mackie, *Ethik. Auf der Suche nach dem Richtigen und dem Falschen*, Stuttgart 1981.

setzt allerdings voraus, daß B nicht nur das Ziel von A, sondern auch die Nebenfolgen seiner Handlungswahl (bei Onkel Karl ist diese Nebenfolge der eigene Tod, bei B handelt es sich um eine Verlängerung seines Gefängnisaufenthaltes) „akzeptieren" kann. Da dies in beiden Fällen nicht der Fall ist, scheint auch hier die Entscheidung in einem normativen Sinne irrational zu sein. B kann nicht sagen: „Es ist zwar rational, was du tun willst (nämlich zu gestehen), aber du solltest nicht gestehen." Äußert er diesen Satz, dann gebraucht er zwar „sollen" normativ, aber „rational" entgegen der konsequentialistischen Auffassung deskriptiv[14]. Aber selbst wenn B sich auf den Standpunkt stellt, A sollte sein Ziel (unbeschadet seiner Nebenfolgen) konsequentialistisch rational verfolgen, ergibt sich eine paradoxe Konsequenz: Da die Situation symmetrisch ist und es sicher zutrifft, daß B seine eigenen Ziele ohnehin akzeptiert, gilt diese Handlungsanweisung a fortiori auch in Selbstanwendung. Man könnte sich vorstellen, daß B zu sich selbst sagt: „Ich sollte gestehen." Zugleich befürwortet er jedoch als rationaler Konsequentialist, daß weder A noch B gestehen.

Deutlicher wird dieses Dilemma einer normativ interpretierten konsequentialistischen Rationalitätskonzeption, wenn man eine dritte Person C auftreten läßt. C ist ebenfalls Konsequentialist und wird A anraten, sein primäres Ziel konsequentialistisch rational zu verfolgen (und das heißt hier: zu gestehen). Aufgrund der Symmetrie der Situation bleibt C nichts anderes übrig, als auch B zu raten zu gestehen. C empfiehlt also jedem von beiden, die Tat zu gestehen. Da C jedoch weiß, daß beide Personen ihr Ziel besser verwirklichen, wenn keiner von beiden ein Geständnis ablegt, empfiehlt er beiden *eine kollektiv irrationale Strategie*. Von beiden gemeinsam befragt, was zu tun sei (beide gestehen oder beide nicht gestehen), würde C antworten: Beide sollten nicht gestehen. Wenn je individuell beiden Personen nur diese beiden Alternativen offenstünden (wenn A gesteht, würde auch B gestehen, und vice versa), wäre es auch *individuell* betrachtet für A und B konsequentialistisch rational, nicht zu gestehen. Da C die Symmetrie der Interaktionsstruktur bekannt ist, weiß er von vornherein, daß ihm nur symmetrische Handlungsempfehlungen offenstehen. Er weiß, daß er nur entweder beiden empfehlen kann, nicht zu gestehen, oder beiden zu gestehen. Mit dieser Prämisse aber müßte C – als rationaler Konsequentialist – beiden empfehlen, nicht zu gestehen. Eine normative Interpretation der konsequentialistischen Rationalitätskonzeption führt unweigerlich in dieses Dilemma. Dies rechtfertigt es, von einer *kollektiven Selbstaufhebung* des Konsequentialismus zu sprechen. *Die kollektive Selbstaufhebung des rationalen Konsequentialismus nimmt bei einer normativen Interpretation des rationalen Konsequentialismus den Charakter immanenter Widersprüchlichkeit an.*

[14] „Deskriptiv" im Sinne von „Du maximierst damit individuell deinen Gefängnisaufenthalt", was ja ein Merkmal der Interaktionssituation des Gefangenendilemmas ist und daher keine neue Information darstellt.

§ 47 Strukturelle Irrationalität

Wenn die konsequentialistische Rationalitätskonzeption mit unserem intuitiven Vorverständnis einer vernünftigen Organisation des Lebens vereinbar wäre, dann könnte der Konsequentialist das Problem der kollektiven Selbstaufhebung als eine minor cura abtun. Wie Thomas Hobbes aufgrund einer pessimistischen und mechanistischen Anthropologie und Jeremy Bentham als ein am öffentlichen Wohl orientierter Sozialreformer könnte er das Koordinationsproblem auf Sanktionen in foro externo verlagern[15]. Er könnte darauf setzen, daß die mit konsequentialistischen Handlungskriterien einhergehenden kollektiv irrationalen Handlungsfolgen durch rechtliche und moralische Normen sowie durch die Ausformung bestimmter Tugenden auf ein erträgliches Maß zu beschränken seien. Der Konsequentialismus hat jedoch über das Problem interpersoneller Koordination hinaus ein strukturgleiches Problem *intrapersoneller Koordination: Eine vernünftige individuelle Strukturierung des Verhaltens ist mit konsequentialistischen Entscheidungskriterien nicht vereinbar.*

Die intrapersonelle Dimension des Koordinationsproblems ist nicht durch eine vollständige Abstraktion von gesellschaftlichen Kontexten gekennzeichnet. Da die Folgen einzelner Handlungen häufig weitgehend von den Entscheidungen anderer Personen abhängen, müssen diese in der Analyse berücksichtigt werden. „Intrapersonell" bezieht sich ausschließlich auf die Beurteilungskriterien selbst. Während das interpersonelle Koordinationsproblem das Auseinanderfallen individueller und kollektiver Optimierung zum Gegenstand hat, geht es beim intrapersonellen Koordinationsproblem um das Auseinanderfallen von Optimierung der Einzelhandlung und Optimierung des Gesamtverhaltens einer Person und zwar auch im gesellschaftlichen Kontext. Wenn man terminologisch beide Probleme erfassen möchte, bietet es sich an, von der *Divergenz punktueller und struktureller Optimierung* zu sprechen. Entsprechend werden wir statt von der intra- und interpersonellen Koordinationsproblematik auch von der *Strukturproblematik individuellen und gesellschaftlichen Handelns* sprechen.

Die Wahl jeder einzelnen einer Folge von Handlungen kann konsequentialistisch rational sein, ohne daß die Folge dieser Handlungen die Wertfunktion maximiert. Diese Thesé klingt paradox, aber sie hat zugleich eine große Verwandtschaft mit einer These, die schon bestätigt wurde, nämlich daß eine Gruppe von Personen zu ihrem Nachteil handeln kann, auch wenn jede einzelne Person dieser Gruppe ausschließlich zu ihrem Vorteil handelt. Ebenso kann es sein, daß jede einzelne Handlung einer

[15] Allerdings ist anzumerken, daß die meisten Theoretiker, die das Schwergewicht ihrer Theorie auf die Regelung der öffentlichen Angelegenheiten legen, Freiraum für individuelle – von moralischen Bindungen weitgehend unbeeinflußte – Lebensgestaltung schaffen wollen, während der rationale Konsequentialist auch bei strikter individueller Befolgung konsequentialistischer Kriterien ein Koordinationsproblem hat.

Person für diese Person vorteilhaft ist, aber sich die Kombination dieser einzelnen jeweils vorteilhaften Handlungen für diese Person nachteilig auswirkt. Im interpersonellen Fall haben wir das Prädikat „kollektiv irrational" für Handlungskombinationen in einer Gruppe gebraucht, die die subjektiven Wertfunktionen keiner der Personen aus dieser Gruppe maximieren – wobei diese Bedingung sicher als hinreichend, aber nicht als notwendig anzusehen ist. Eine kollektiv irrationale Handlungskombination kann ausschließlich aus konsequentialistisch rationalen Handlungen bestehen. Im intrapersonellen Falle sprechen wir analog von der *strukturellen Irrationalität* einer Handlungskombination (bzw. eines Verhaltensmusters). Die These kann daher auch so formuliert werden: *Ein Verhaltensmuster, das sich ausschließlich aus punktuell konsequentialistisch rationalen Handlungen zusammensetzt, kann dennoch strukturell irrational sein.*

Die Gegenüberstellung punktueller und struktureller Optimierung darf nicht mit der Gegenüberstellung von *kurzfristiger* und *langfristiger* Optimierung vermengt werden. So wird heute in der Ökologiediskussion auch dort vor den Schwächen einer kurzfristigen Optimierung gewarnt, wo das angesprochene Problem ausschließlich Folgen der *strukturSprengenden Wirkung punktueller Optimierung* ist und, genau besehen, mit *kurzfristiger* Optimierung nichts zu tun hat. Daß es sich bei der Gegenüberstellung „punktuell/strukturell" und „kurzfristig/langfristig" um ganz heterogene Begriffspaare handelt, kann man sich leicht klarmachen. Das Problem der strukturSprengenden Wirkung konsequentialistischer (d. h. punktueller) Optimierung ist streng von der Frage zu unterscheiden, ob die konsequentialistisch geforderte Zeit-Homogenität der subjektiven Werte adäquat ist. Es erscheint auf den ersten Blick paradox: *Die konsequentialistischen Handlungsnormen haben sowohl einen (unangemessen) langfristigen als auch einen (unangemessen) punktuellen Charakter.* Das erste Merkmal ist für eine Kritik des konsequentialistischen Idealtypus rationalen Verhaltens eher unergiebig, während das zweite Merkmal die *Defizite des Konsequentialismus im Bereich der interpersonellen Koordination als Sonderfall der Inadäquanz des Konsequentialismus angesichts struktureller Aspekte des Handelns generell aufweist.*

§ 48 Selbstkontrolle als Orientierung an struktureller Rationalität

Strukturelle Irrationalität individueller Verhaltensmuster läßt sich nur durch eine spezifische Form der Selbstkontrolle beheben. Zur Illustration wählen wir zunächst ein alltägliches, sehr einfach strukturiertes Beispiel. Ein langjähriger Raucher stellt sich die Frage, ob er mit dem Rauchen aufhören soll. Die subjektiven Bewertungen des Rauchers sind dabei dergestalt, daß sich das Zeitintegral seiner Wertfunktion monoton erhöht, je früher er mit dem Rauchen aufhört. In dieser Bewertung geht nicht nur der allgemeine erwartete Gesundheitszustand des Rauchers ein (eine pro-

babilistische Abschätzung), sondern auch der Genuß, den er zweifelsfrei beim Rauchen einer Zigarette empfindet, möglicherweise auch der Schaden, den er anderen Personen zufügt etc. Wenn er dennoch nicht mit dem Rauchen aufhört, so scheint es sich um einen eindeutigen Fall von akrasia oder Willensschwäche zu handeln: Er weiß, was bei Abwägung aller Wertungsaspekte am günstigsten wäre und kann sich dennoch nicht entschließen, es zu tun. Diese Interpretation ist jedoch keineswegs zwingend. Es könnte sein, daß der Raucher *konsequentialistisch rational* handelt, wenn er mit dem Rauchen nicht aufhört, obwohl gilt, daß sich das Zeitintegral seiner Wertfunktion monoton erhöht, je früher er mit dem Rauchen aufhört. Denn unter normalen Umständen steht er als Konsequentialist zu keinem Zeitpunkt vor der Entscheidung, das Rauchen aufzugeben. „Unter normalen Umständen" insofern, als wir annehmen wollen, daß ihm keine Möglichkeiten offenstehen, sich selbst zu binden, sei es über den Umweg einer (Selbst-)Sanktionierung des Rauchens oder indem er sich einem physischen Zwang unterwirft. Normalerweise steht der Raucher, der sich bewußt ist, daß frühzeitiges Aufhören seine subjektiven Werte maximiert, vor der Frage, ob er an einem bestimmten Tag rauchen soll oder nicht – ja eigentlich nur, ob er sich die nächste Zigarette anzünden soll oder nicht.

Betrachten wir diese *punktuellen* Alternativen etwas genauer: Wie immer bei konsequentialistischer Abwägung gilt es zunächst, die kausalen Folgen bzw. die Wahrscheinlichkeitsverteilung zukünftiger Weltverläufe abzuschätzen und zu bewerten. Der Rauch-Verzicht am heutigen Tage bringt einen gewissen, wenn auch geringen gesundheitlichen Nutzen. Dafür wird sich der Raucher an diesem Tag sehr quälen – gerade der erste Tag ist ja erfahrungsgemäß besonders belastend. Natürlich müssen wir eine langfristige Abwägung vornehmen: Entscheidend sind nicht die augenblicklichen gesundheitlichen Vorteile, sondern die langfristigen. So mag es sein, daß sich der Gesundheitszustand des starken Rauchers am ersten Tag des Verzichtes deutlich verschlechtert, der eine Tag des Rauchverzichtes jedoch selbst dann gesundheitlich von Nutzen ist, wenn er das Rauchen am nächsten Tag wieder fortsetzt. Möglicherweise ergeben sich aus seiner schlechten Stimmung gravierende Weiterungen im privaten Bereich: Streit, unbedachte Äußerungen, Verletzungen. Auch hier dürfen wir allerdings nicht nur die Vorgänge dieses einen Tages betrachten, sondern die langfristigen Auswirkungen. Der Streit könnte mit einer gewissen Wahrscheinlichkeit eine dauerhafte Verstimmung auslösen, aber vielleicht auch alte und verdrängte Konflikte bereinigen. Die Details des Dramas brauchen uns nicht zu interessieren, entscheidend ist die Antwort auf folgende Frage: Könnte es sein, daß der Saldo dieser (L) konsequentialistischen Abschätzung negativ ausfällt, und dennoch die Prämisse aufrechterhalten werden kann, daß es für den Raucher desto günstiger ist, je eher er mit dem Rauchen aufhört? – Die Antwort lautet „ja".

Um dies einzusehen, genügt es, das konsequentialistische Entscheidungskriterium uneingeschränkt anzuwenden. Es liegt auf der Hand, daß die langfristigen Folgen dieses *einen* Tages des Rauchverzichtes davon abhängen, was danach geschieht. Angenommen, am nächsten Tag wird das Rauchen fortgesetzt. Dann kann dieser eine Tag, auch bei langfristiger Betrachtung, mehr Verdruß gebracht haben als gesundheitliche Vorteile. Die Formulierung „bei langfristiger Betrachtung" darf dabei nicht mißverstanden werden: Es geht um die langfristigen Folgen des Rauchverzichtes an *einem bestimmten* Tag und nicht um die Abwägung der Vor- und Nachteile eines langfristigen Rauchverzichtes. Wenn es jedoch beim Rauchverzicht bleibt, dann kann es ebenfalls sein, daß die gesundheitlichen Vorteile dieses einen Tages der Verkürzung der Rauchphase die Nachteile, die gerade dieser erste Tag des Rauchverzichts verursacht, nicht aufwiegen. Dazwischen gibt es eine Vielzahl weiterer Möglichkeiten – vorübergehender Rauchverzicht von einem Monat etc. – für die sich keine wesentlich neuen Aspekte mehr ergeben. Dieses Argument macht implizit vom Dominanz-Prinzip Gebrauch, dessen Anwendung voraussetzt, daß die Wahrscheinlichkeiten der für die Bewertung der Handlungskonsequenzen relevanten Umstände handlungsunabhängig sind. Wir können diese Prämisse, die die Selbstrechtfertigung des vergeblich gegen seine Abhängigkeit ankämpfenden Rauchers ohnehin prägt, explizit machen: Ob ich heute rauche oder nicht, spielt für die Wahrscheinlichkeit, daß ich irgendwann mit dem Rauchen ganz aufhöre, keine Rolle. Unter dieser Bedingung ist die Anwendung des Dominanzprinzips zulässig, und damit ist es – punktuell betrachtet – an jedem Tag konsequentialistisch rational weiterzurauchen.

Wenn man jedoch nur folgende Alternativen betrachtet: Der Raucher hört ab heute für immer mit dem Rauchen auf, der Raucher hört ab morgen für immer mit dem Rauchen auf, der Raucher hört ab übermorgen für immer mit dem Rauchen auf etc., dann kann die erste Alternative dennoch die besten Folgen haben. Das Problem des Konsequentialisten besteht darin, daß diese Alternativen i. allg. nicht Gegenstand der Entscheidung sind: Der rationale Konsequentialist soll *jede einzelne Entscheidung* so treffen, daß ihre Folgen optimal sind. Unbezweifelbar steht dem Raucher an jedem einzelnen Tag die Entscheidung offen, ob er an diesem Tag raucht oder nicht. Ja es steht ihm bei jeder einzelnen Zigarette die Entscheidung offen, diese anzuzünden oder nicht – es sei denn, dieser Vorgang ist schon so stark dispositionell determiniert, daß er nicht mehr als eine (absichtliche) Handlung gelten kann. Nicht jede Folge punktuell konsequentialistisch rationaler Einzelentscheidungen aber ist strukturell betrachtet rational: Es gibt unter Umständen ein anderes Verhaltensmuster, das das Zeitintegral der Wertfunktion gegenüber dieser Folge punktuell konsequentialistisch rationaler Einzelentscheidungen erhöhen würde. Im Raucher-Beispiel wären diese strukturell besseren, aber sich nicht notwendigerweise aus durchgängig konsequentialistisch rationalen Einzelent-

scheidungen zusammensetzenden Verhaltensmuster: 1.) der Raucher hört ab heute für immer mit dem Rauchen auf, 2.) der Raucher hört ab morgen für immer mit dem Rauchen auf, 3.) der Raucher hört ab übermorgen für immer mit dem Rauchen auf etc. – in dieser Reihenfolge.

Selbstkontrolle kann darin bestehen, seine Handlungen je punktuell so zu wählen, daß sie sich in eine wünschenswerte Struktur des eigenen Handelns einfügen. Selbstkontrolle ist in diesem Sinne nichts anderes als eine Orientierung an struktureller Rationalität, ohne daß sich an den jeweils vorhandenen Handlungsspielräumen etwas änderte. Eine andere Möglichkeit besteht darin, die vorhandenen Handlungsspielräume einzugrenzen oder zu modifizieren[16]. Aber auch diese indirekte (Meta-)Strategie orientiert sich an struktureller Rationalität.

Selbstkontrolle muß nicht ausschließlich der Eingrenzung strukturSprengender Auswirkungen konsequentialistisch rationalen Handelns gelten – in vielen Fällen nimmt sie eine andere Funktion wahr. Eine dieser Funktionen kann man als Wahrung der Einheitlichkeit der Person bezeichnen. Man vermutet, daß sich in einer bestimmten zukünftigen Situation die eigenen Präferenzen (Nutzen, Wertfunktionen) ändern könnten. Eine große zeitliche Varianz der eigenen subjektiven Wertungen aber gefährdet langfristige Projekte. Man trifft Vorkehrungen, die entweder diese Veränderungen der subjektiven Wertungen verhindern oder aber ausschließen, daß sich diese Veränderungen in entsprechenden Handlungen niederschlagen – unabhängig davon, ob man für sein eigenes Handeln konsequentialistische Kriterien zugrundelegt. Die Entscheidung Nastassja Filippownas, Fürst Myschkin (in Dostojewskis *Idiot*) nicht mehr zu sehen, ist ein Beispiel für den ersten Fall, während Odysseus, der sich von seinen Gefährten an den Segelmast binden läßt, um die Sirenen hören zu können, ein Beispiel für den zweiten Fall ist. Eine andere Funktion ist die der Vermeidung irrationalen Handelns. Man befürchtet, sich in einer zu erwartenden Situation irrational – speziell willensschwach – zu verhalten. Die subjektiven Bewertungen bleiben dabei unverändert. Man kann Vorkehrungen treffen, daß diese Situation erst gar nicht eintritt oder sie z. B. durch selbstgewählte Sanktionen so modifizieren, daß es nicht zu dem befürchteten Verhalten kommt.

Wenn der Raucher einen Freund beauftragt, dafür zu sorgen, daß er nicht mehr an Zigaretten kommt, grenzt er einerseits Handlungsoptionen

[16] Die unterschiedlichen Möglichkeiten und Techniken individueller und kollektiver *Selbstbindung* – nur eine aus der Vielfalt der unterschiedlichen Möglichkeiten von Selbstkontrolle – werden in immer neuen Facetten von J. Elster dargestellt. Jedoch besteht ein gravierender Unterschied: Elster diskutiert Selbstbindung unter dem Aspekt der Vermeidung *irrationalen* Handelns, während es hier um die Vermeidung der strukturSprengenden Wirkungen *konsequentialistisch rationalen* Handelns geht. Vgl. J. Elster *Ulysses and the Sirens. Studies in Rationality and Irrationality*, Cambridge u. a. 1979 sowie *Sour Grapes. Studies in the Subversion of Rationality*. Cambridge 1983, teilweise wiedergegeben in: *Subversion der Rationalität*, Frankfurt/New York 1987.

aus, andererseits kommt eine neue Handlungsoption hinzu, nämlich sich selbst die Möglichkeit des Rauchens zu nehmen. Die Entscheidung für diese Handlungsoption ist gerade dann punktuell konsequentialistisch rational, wenn er davon ausgehen muß, daß er sich in allen Situationen punktuell konsequentialistisch rational verhält. Der rationale Konsequentialist beraubt sich selbst bestimmter Entscheidungsoptionen, gerade weil er weiß, daß er punktuell konsequentialistisch rational handelt. Die Kantische These, eine teleologische Ethik sei mit der Freiheit der Person nicht vereinbar, erfährt auf diese Weise eine neue und in dieser Form unerwartete Bestätigung. Dies darf jedoch nicht dazu verleiten, den rationalen Konsequentialisten für eine willensschwache Person zu halten – das ist er per definitionem gerade nicht: Der rationale Konsequentialist tut immer das, was angesichts der jeweils gegebenen Wertfunktin optimal ist: Er erfüllt alle Anforderungen der konsequentialistischen Theorie.

Das Raucherbeispiel hat den Vorteil, in besonders reiner Form das Strukturproblem individuellen Handelns im Rahmen des Konsequentialismus zu verdeutlichen. Es hat den Nachteil, daß die recht komplexen Annahmen intuitiv nicht sofort einleuchten. Daher sei noch ein zweites, kürzeres Beispiel genannt. Im allgemeinen sind Drohungen (oder allgemeiner, Ankündigungen bestimmter Handlungen, die einem anderen Akteur zum Nachteil gereichen und ihn zu einem kooperativen Verhalten veranlassen sollen), soweit sie von einer konsequentialistisch rationalen Person geäußert werden, unglaubwürdig. Es ist ein typisches Charakteristikum von Drohungen, daß ihre Verwirklichung dann, wenn die Bedingungen dafür gegeben sind, nicht mehr konsequentialistisch rational ist. Wäre es konsequentialistisch rational, eine Drohung zu verwirklichen, dann würde ein wohlinformierter Akteur (der für sich und andere konsequentialistisch rationales Verhalten annimmt) ohnehin damit rechnen, so daß die Äußerung der Drohung überflüssig ist. Wenn wir annehmen, daß es Situationen gibt, in denen die Glaubwürdigkeit einer Drohung vorteilhaft ist, dann hat der rationale Konsequentialist ein Interesse daran, nicht für einen solchen gehalten zu werden. Und wenn wir zusätzlich annehmen, daß das Geheimhalten der eigenen Handlungsorientierung schwierig und mit großen Belastungen verbunden ist, hat der rationale Konsequentialist ein Interesse daran, kein rationaler Konsequentialist zu sein. In diesem Beispiel spielt die Interaktion – anders als beim Raucherbeispiel – eine Rolle. Dennoch sind wir auch mit diesem Beispiel im Rahmen einer Analyse des intrapersonellen Koordinationsproblems geblieben; denn wir haben uns ausschließlich auf die Frage konzentriert, ob eine konsequentialistisch rationale Person je individuell ihre Ziele verwirklicht. Die Antwort lautet, „nein" – und dies ist ein starkes Argument gegen die konsequentialistische Rationalitätskonzeption.

15. Kapitel: Handlungsgründe

§ 49 *Deskriptive und normative Begründungen*

Fragen wie: „Sollte ich dies oder jenes tun?", oder „Warum hast du das getan?" sind Fragen nach Handlungsgründen: Welche Gründe sprechen dafür, eher dies als jenes zu tun? Welche Gründe rechtfertigen es, das getan zu haben? Die mit Fragen dieser Art korrespondierende *philosophische Frage* lautet: Was kann als Grund dafür gelten, eine Handlung zu vollziehen?

Ein Beispiel:

A: Warum hast du deine Bank gewechselt?

B: Weil mir von der neuen Bank höhere Zinsen eingeräumt werden.

A: Warum willst du höhere Zinsen?

B: Weil ich mein Einkommen aufbessern will.

A: Warum willst du dein Einkommen aufbessern?

B: (Es ist anzunehmen, daß B spätestens jetzt ungeduldig wird und antwortet): Weil ich es eben will! Oder: Das versteht sich doch von selbst. Aber auch wenn B geduldig bleibt, wird er A irgendwann keine rechte Antwort auf seine Warum-Fragen mehr geben können. Die Begründungskette ist an ihr Ende gelangt. Dieses Ende besteht in diesem Fall im Hinweis auf ein *persönliches Interesse*. Doch das muß nicht so sein.

Vielleicht antwortet B auf die letzte Frage: „Ich möchte für meine Kinder eine Rücklage einrichten, damit sie später leichter zu ökonomischer Selbständigkeit finden." Man kann dieses Motiv nicht ohne weiteres als ein „persönliches Interesse" von B bezeichnen – es sei denn, daß B einen Vorteil davon hat, wenn seine Kinder frühzeitig selbständig werden. Wenn B auf die letzte Frage antwortet: „Ich möchte dieses Geld dem Roten Kreuz spenden", dann entspricht dieser Handlungsgrund unter normalen Umständen sicherlich nicht einem persönlichen Interesse von B. Dennoch kann diese Aussage eine taugliche Begründung für seine Entscheidung sein.

Für A sind die Antworten von B auch eine Art von *Erklärung* für Bs Entscheidung, die Bank zu wechseln (im folgenden Entscheidung/Handlung E genannt). Wenn er schon vor dieser Entscheidung gewußt hätte, daß B sich Gedanken macht, wie er sein Einkommen erhöhen kann, hätte A möglicherweise *prognostiziert*, daß B die Bank wechseln wird. Zumindest ist die Apriori-Wahrscheinlichkeit, die A der Entscheidung E zumißt, niedriger als die Wahrscheinlichkeit, die A aufgrund dieser zusätzlichen Information E zumessen würde. Die Erklärung von E durch Angabe von subjektiven Handlungszielen gilt A i. allg. jedoch nicht als eine *kausale Erklärung*, da A weiß, daß B nicht durch Naturgesetze dazu gezwungen ist,

das zu tun, was seinen Zielen optimal entspricht. Mit anderen Worten: A nimmt an, daß B auch hätte anders handeln können. In welchem Sinne akzeptiert A die Antworten von B als Begründung von E? Daß die Antworten eine Erklärung von E abgeben, ist jedenfalls keine hinreichende Bedingung dafür, daß sie zugleich eine Begründung von E sind.

Wenn B die in Anbetracht seiner subjektiven Ziele optimale Entscheidung getroffen hat, ist A nicht gezwungen zuzugeben, daß B zu dieser neuen Bank wechseln sollte. Denn das würde bedeuten, wie in § 46 gezeigt wurde, die Handlungsziele von B bei Berücksichtigung aller Nebenfolgen und Wertaspekte zu akzeptieren: A müßte damit einverstanden sein, daß B die von ihm genannten Ziele erreicht. In einer Situation wie der oben genannten wird das in den meisten Fällen der Fall sein. Dieser Umstand darf jedoch nicht darüber hinwegtäuschen, daß mit der Äußerung eines normativen (abgetrennten) Sollenssatzes zusätzliche Annahmen einfließen, die sich nicht nur auf die subjektiven Handlungsziele von B beziehen. Ohne solche zusätzlichen Annahmen kann die Angabe persönlicher Handlungsziele nur subjektive Gründe liefern, d. h. Gründe, die es dem Hörer erleichtern zu verstehen (oder auch: zu erklären), warum B sich so und nicht anders entschieden hat. Diesem Typus der Begründung von Handlungen und Entscheidungen (im folgenden *subjektive Begründung* genannt) entspricht die *deskriptive* Verwendung von Sollenssätzen, wie sie in § 46 erläutert wurde.

Der andere Typus der Begründung (im folgenden *objektive Begründung* genannt) korrespondiert mit der normativen Verwendung von Sollenssätzen. Eine gelungene Begründung der zweiten Art läßt den Hörer A sagen: B sollte E tun. Dabei ist es nicht erforderlich, daß die subjektiven Handlungsziele von B moralischer Natur sind. Auch eine objektive Begründung kann sich auf subjektive Handlungsziele beziehen.

Den Unterschied zwischen subjektiver (deskriptiver) und objektiver (normativer) Begründung von Handlungen und Entscheidungen kann man sich auch folgendermaßen klarmachen: Eine deskriptive Begründung erlaubt nur *in der ersten Person* die Ableitung von (abgetrennten) Sollenssätzen. Man stelle sich vor, zu sich selbst zu sprechen: „Ich habe das ausschließliche Ziel Z, und H ist die einzige Möglichkeit, Z zu erreichen, daher sollte ich H tun." Eine normative Begründung erlaubt dagegen auch in der dritten Person die Ableitung eines (abgetrennten) Sollensatzes: „Er hat diese Wünsche; er sollte sich diese Wünsche erfüllen – er sollte H tun." Man beachte, daß diese Interpretation mit der Theorie des naturalistischen Fehlschlusses vereinbar ist.

§ 50 Nicht-konsequentialistische Handlungsgründe

Ein zweites Beispiel:

A: Warum hast du an dieser Demonstration teilgenommen?

B: Weil ich gegen die weitere Aufrüstung mit nuklearen Mittelstreckenwaffen bin.

A: Warum bist du gegen die weitere Aufrüstung mit Mittelstreckenwaffen?

B: Weil diese einen destabilisierenden Effekt haben, wenn es zu einer internationalen Krise kommen sollte.

A: Was hast du dagegen einzuwenden?

B: Ich möchte, daß der Frieden sicherer, nicht unsicherer wird.

Eine Gemeinsamkeit zwischen den beiden Beispielen ist, daß B am Ende seine persönliche Einstellung als Grund angibt. Daher kann auch in diesem Beispiel in gleicher Weise eine objektive (normative) und eine subjektive (deskriptive) Begründung unterschieden werden: Die letzte Antwort von B stellt jeweils einen Rückzug auf subjektive Gründe dar. Nur dann, wenn auch der Hörer dieses Handlungsziel akzeptiert, kann er die Antworten als eine (objektive) Begründung der Entscheidung akzeptieren[1].

Ein Unterschied zwischen dem ersten und dem zweiten Beispiel besteht darin, daß B im zweiten Beispiel moralisch motiviert ist und im ersten nicht. Dieser Unterschied soll uns aber hier nicht weiter beschäftigen. Es gibt einen anderen Unterschied, der sich auf die logische *Beziehung von Handlungsziel und Handlung* bezieht. Im ersten Beispiel ist die Handlung fraglos als Mittel geeignet (oder zumindest gedacht), um das Ziel zu erreichen. Im zweiten Beispiel ist es recht unwahrscheinlich, daß B die Abrüstung der Mittelstreckenwaffen mit seiner Beteiligung an der Friedensdemonstration erreichen kann. Ja vermutlich erhöht sich dadurch nicht einmal die Wahrscheinlichkeit der Abrüstung. Dennoch ist diese *Form* der Begründung einer Handlung nicht ungewöhnlich: „Warum fährst du nicht mit dem Wagen in die Stadt?" – „Weil ich meine, die Innenstadt sollte von Privatverkehr freigehalten werden." Auch wenn der Verzicht, das eigene Fahrzeug zu benutzen, die Innenstadt nicht von Privatverkehr freihält, scheint diese Antwort eine denkbar Begründung zu sein.

Die philosophischen Antworten auf die Frage, wie eine Entscheidung/Handlung begründet werden kann, lassen sich danach unterscheiden, welche Aussagen sie als Begründungen von Handlungen zulassen, aber auch danach, welche Beziehung zwischen diesen Aussagen und der Handlung angenommen wird. Wenn eine Handlung damit begründet wird, daß sie ein geeignetes Mittel ist, die Ziele der handelnden Person zu erreichen, dann handelt es sich um eine *konsequentialistische Begründung*. Das erste

[1] „Akzeptieren" hier immer in dem schon in § 46 gebrauchten schwachen Sinne von „einverstanden sein, daß der Akteur seine Ziele erreicht".

Beispiel beinhaltet eine konsequentialistische Begründung, während die Begründung des zweiten Beispiels *keinen konsequentialistischen Charakter* hat. Subjektive Handlungsgründe sind nach konsequentialistischem Verständnis immer motivierende Absichten, welche sich auf bestimmte kausale Folgen der Handlung beziehen. Diese kausalen Folgen sind weder Regeln noch Handlungen, sondern Zustände der Welt („state of affairs") im Sinne von § 4.

Betrachten wir den klassischen Utilitarismus. Nach dieser Theorie ist eine Handlung genau dann wohlbegründet, wenn sie die Nutzensumme maximiert. Dies ist ein objektives (nicht subjekt-zentriertes) Kriterium. Zugleich wird jedoch von dieser Theorie die ideale moralische Person definiert. Die motivierenden Absichten der idealen moralischen Person im Sinne des klassischen Utilitarismus sind auf die Maximierung der Nutzensumme gerichtet. Die Theorie des Rechten überträgt sich auf die Theorie legitimer Handlungsgründe, sie schränkt die Menge moralisch zulässiger subjektiver Handlungsgründe ein.

Inwieweit auch konsequentialistische Theorien ein bestimmtes Maß an Universalisierbarkeit enthalten müssen, ist umstritten. In § 46 wurde dargelegt, inwiefern und in welchem Sinne auch konsequentialistische Handlungsgründe unter dem logischen Zwang der Allgemeingültigkeit stehen, wenn sie normativ (und nicht deskriptiv) verstanden werden. Im Grenzfall läßt ein hoher Grad der Universalisierung ausschließlich strikt-konsequentialistische Handlungsgründe zu[2]. Abschwächungen der Universalisierbarkeitsforderung erlauben Handlungsgründe, die sich im anderen Grenzfall auch auf beliebige subjektive Präferenzen beziehen können[3], soweit sie nur die Konsistenzbedingungen der Theorie T1 aus § 8 erfüllen.

In den vorangehenden Kapiteln wurde gezeigt, daß eine konsequentialistische Auffassung praktischer Rationalität aus einer Reihe von Gründen inadäquat ist. Diese Analyse bedarf jedoch noch der Abrundung. Denn wenn Handlungsgründe grundsätzlich einen konsequentialistischen Charakter hätten, dann wäre eine nicht-konsequentialistische Theorie praktischer Rationalität kaum aufrechtzuerhalten – es sei denn, man ist bereit, auf das Prinzip des Konstruktivismus (s. § 42) zu verzichten. Wenn normative Theorien immer auch im Sinne eines Entscheidungsverfahrens zu verstehen sind, dann muß es neben konsequentialistischen Handlungsgründen noch Handlungsgründe anderer Art geben.

B gibt im zweiten Beispiel keine konsequentialistische Begründung für sein Handeln, da sich die motivierenden Absichten von B nicht auf die kausalen Folgen seines Handelns beziehen. Die kausalen Folgen seines Handelns sind nicht Bestandteil seiner Handlungsziele. Sind nicht-konse-

[2] R. M. Hare, *Freedom and Reason*, Oxford 1963, Kap. 2 und *Moralisches Denken*, Frankfurt/M. 1992, ist sogar der Auffassung, daß die Universalisierbarkeitsregeln der Moralsprache eine handlungsutilitaristische Ethik logisch erzwingen.

[3] Dem entspricht die sog. „normative Interpretation" der Entscheidungtheorie.

quentialistische Gründe, wie die des zweiten Beispiels, nur Scheingründe, die auf einer laxen Redeweise beruhen, oder gibt es neben dem konsequentialistischen andere Modelle der Handlungsbegründung?

Für das, was wir als „konsequentialistisches Begründungsmodell" bezeichnet haben, sind auch andere Begriffe gebräuchlich, wie z. B. die des „intentionalen" und des „teleologischen Begründungsmodells". Aus Gründen, die schon dargelegt wurden[4], sollte jedoch zwischen „teleologisch" und „konsequentialistisch" unterschieden werden: „intentional" scheidet deshalb aus, weil damit der Eindruck erweckt wird, andere Erklärungsmodelle bezögen sich auf nicht-intentionales Handeln. G. H. v. Wright war lange Zeit einer der exponiertesten Vertreter einer Theorie konsequentialistischer Handlungsgründe mit universalem Anspruch[5]. Seit Ende der 70er Jahre räumt v. Wright jedoch auch nicht-konsequentialistischen Begründungs- bzw. Erklärungsmodellen einen gleichberechtigten Stellenwert ein. Wir knüpfen im folgenden an seine Überlegungen an, obwohl wir weder die Terminologie noch die spezifische Interpretation im Sinne „innerer" und „äußerer" Determinanten übernehmen können[6].

Wir betrachten zunächst den Fall der Begründung „in der ersten Person", also Begründungen, die nur subjektive Geltung beanspruchen, und unterscheiden drei Formen subjektiver Begründung:

1) Ich streiche mir ein Brot, weil ich Hunger habe.

2) Ich schließe das Fenster, weil mich meine Freundin darum gebeten hat.

3) Ich beteilige mich an einer Radtour, weil ich es versprochen habe.

Die hinter „weil" jeweils angeführte Proposition begründet die betreffende Handlung.

(1) paßt sich gut in das konsequentialistische Begründungsmodell ein: Mein Ziel ist es, einen Zustand zu erreichen, in dem mein Hunger gestillt ist. Das Butterbrot zu streichen und anschließend zu verzehren, ist dazu ein geeignetes Mittel. Wenn ich allerdings eine Diätkur mache und fest entschlossen bin, das Butterbrot nicht zu verzehren, verkompliziert sich die Analyse. Hier würde man (1) wohl zunächst als eine kausale Erklärungsskizze und nicht als Begründung ansehen.

(2) und (3) lassen sich dagegen unter normalen Umständen kaum als konsequentialistische Begründungen ansehen. In beiden Fällen müßte es ein subjektives Ziel geben, das durch die gewählte Handlung erreicht wird. Welches Ziel verfolge ich mit dem Schließen des Fensters: Will ich damit vermeiden, daß Mißstimmung aufkommt, oder möchte ich meiner

[4] Vgl. § 19 und § 25.

[5] Allerdings unter dem Etikett der teleologischen bzw. intentionalen Handlungserklärung. Vgl. G. H. v. Wright, *The Varieties of Goodness*, London 1963; leicht abgeschwächt vertritt v. Wright diesen Standpunkt auch noch in *Explanation and Understanding* (1971), dt. *Erklären und Verstehen*, Königstein/Ts. ²1984, Kap. III.

[6] Vgl. G. H. v. Wright, „Das menschliche Handeln im Lichte seiner Ursachen und Gründe", in: Lenk (1979), Bd. 2/II.

Freundin eine Freude machen? Selbstverständlich können solche Handlungsziele im Einzelfall eine Rolle spielen, aber unter gewöhnlichen Umständen ist meine Handlung nicht dadurch motiviert, daß ich Ziele dieser Art anstrebe. Selbst die Formulierung: „Ich schließe das Fenster, weil ich einer Bitte meiner Freundin nachkommen möchte", wirkt unnatürlich. Mein Ziel ist es nicht, einer Bitte nachzukommen, sondern ich komme einer Bitte nach, indem ich das Fenster schließe, und ich schließe es, weil ich darum gebeten worden bin. Es reicht als Begründung der Handlung i.allg. aus, darauf zu verweisen, daß ich gebeten wurde, das Fenster zu schließen. Mit v. Wright (1979) nennen wir Begründungen dieser Art *„imperativisch"*.

Die Begründung in (3) ist dagegen von anderer Natur. Zwar ist es auch hier wenig plausibel, nach einem Handlungsziel zu suchen, das durch diese Handlung erreicht wird, auch wenn es Fälle gibt, in denen das angemessen wäre. Es könnte z. B. sein, daß ich nur deshalb trotz des schlechten Wetters eine Radtour unternehme, weil ich befürchte, daß meine Bekannten sonst verärgert wären. Sie wären deshalb verärgert, weil ich zuvor versprochen hatte zu kommen. Insofern wäre es in diesem Fall zutreffend zu sagen: „Ich komme, weil ich es versprochen habe." Die Tatsache, daß ich es versprochen habe, ist aber dann nicht der eigentliche Grund für mein Kommen, sondern mein Wunsch, eine Verärgerung meiner Bekannten zu vermeiden. Das Versprechen spielt dabei nur die Rolle einer (kausalen) Determinante dieser Verärgerung. Normalerweise aber ist die Tatsache, daß ich mein Kommen versprochen habe, unmittelbarer Grund für mein Kommen. Natürlich kann es andere Gesichtspunkte geben, die diesen Grund nicht wirksam werden lassen, aber allein die Tatsache, daß ich x versprochen habe, ist ein Grund, x auszuführen. Mit v. Wright (1979) nennen wir diesen zweiten Typus nicht-konsequentialistischer Begründungen *„normativisch"*.

Eine konsequentialistische Theorie praktischer Rationalität wäre nur dann adäquat, wenn sich jede adäquate Handlungsbegründung im konsequentialistischen Rahmen rekonstruieren ließe. Offensichtlich spielen jedoch bei subjektiven Begründungen Handlungsgründe eine Rolle, die sich konsequentialistisch nicht rekonstruieren lassen.

Zu prüfen bleibt, ob sich dieser Befund auf subjektive (deskriptive) Begründungen beschränkt. Bei konsequentialistischen Begründungen ist der Übergang von einer subjektiven (deskriptiven) zu einer objektiven (normativen) Begründung dadurch gekennzeichnet, daß die Handlungsziele (vom jeweiligen Hörer) akzeptiert werden können. Bei nicht-konsequentialistischen Begründungen gibt es etwas Analoges: Bei (2) wird der Hörer dann zu dem (abgetrennten) Sollensurteil kommen, „er sollte das Fenster schließen", wenn er es (ebenso wie der Sprecher) für richtig hält, eine Bitte dieser Art zu befolgen. Bei (3) gilt gleiches für die Institution des Versprechens. Gegenüber den subjektiven Modellen der Begründung

ändert sich daher insofern nichts: Auch bei normativer Betrachtungsweise gibt es Begründungen, die sich konsequentialistisch nicht rekonstruieren lassen.

Man beachte, daß bei *objektiven* Begründungen, unabhängig davon, ob sie auf konsequentialen[7], auf imperativischen oder normativischen Handlungsgründen beruhen, eine Form der *Verallgemeinerung* eine Rolle spielt. Eine objektive (normative) Begründung folgt dem Muster einer *ersten Charakterisierung der konkreten Handlung*, ihrer *Unterordnung* unter eine allgemeinere Handlungsbeschreibung und schließlich (wenn der Begründungsprozeß zum Abschluß kommt) einer *Befürwortung eines Handlungskriteriums* – was nichts anderes ist als die Befürwortung eines Handlungstypus. Derjenige, der eine Handlung gegenüber kritischen Hörern begründet, verallgemeinert die Beschreibung seiner Handlung und sucht Zustimmung zur Handlung *unter einer bestimmten Beschreibung* zu erreichen. Welche Beschreibungen in Frage kommen, hängt in hohem Maße von dem Komplex der handlungsleitenden, motivierenden und vorausgehenden Absichten und den Umständen der Handlung ab. Der Unterschied zwischen konsequentialistischen und nicht-konsequentialistischen Begründungen schlägt sich im Charakter desjenigen Handlungstypus nieder, bei dem die Begründungsketten jeweils enden.

Eine Handlung ist nach Auffassung einer konsequentialistischen ethischen Theorie T genau dann wohlbegründet, wenn man sie als erwartungswertmaximierend bezüglich der von der Theorie T postulierten Wertfunktion über Zustände (Weltverläufe) beschreiben kann. Erwartungswertmaximierend zu sein ist ein Merkmal von Handlungen. Kant hat mit dem moralischen Gesetz ein anderes begründendes Merkmal von Handlungen benannt: Die normative Begründung endet dann, wenn eine Handlung als unter dem moralischen Gesetz stehend beschrieben ist. Es gibt eine wichtige Parallelität beider Theorien: Beide Begründungsmodelle beruhen darauf, daß ein Handlungstypus als begründend ausgezeichnet wird: Nur diejenigen Handlungen, die unter diesen Typus fallen, sind wohlbegründet. Die Auszeichnung dieses Handlungstypus hat jeweils einen rechtfertigenden Status: Handlungen, die diesem Typ angehören, sind gerechtfertigt. Daraus darf aber nicht (im Geist der klassischen Auffassung axiomatischer Theorien) geschlossen werden, daß diese Auszeichnung nicht selbst wiederum der Kritik unterzogen werden kann. Einige Elemente der Kritik des konsequentialistischen Begründungsmodells wurden in diesem und im vorausgegangenen Kapitel dargelegt.

Es macht die Stärke von Konsequentialismus und Kantianismus aus, daß sie jeweils einen einzigen begründenden Handlungstypus benennen und charakterisieren können, während die Alltagssprache und die unser tägliches Handeln strukturierenden normativen Institutionen eine Vielfalt

[7] Konsequentiale Gründe sind diejenigen, die nach konsequentialistischer Auffassung allein geeignet sind, eine Handlung zu begründen.

begründender Handlungstypen kennen. Beide Theorien reduzieren diese Vielfalt auf ein einziges einfaches Kriterium, und beide Theorien haben darin auch ihre besondere Schwäche. Die Kritik des Konsequentialismus ist nicht zugleich eine Rechtfertigung seines ideengeschichtlichen Antipoden – dazu haben Kantianismus und Konsequentialismus unter einem systematischen Gesichtspunkt zu viele Gemeinsamkeiten.

§ 51 Kohärenz ohne Konsequentialismus

In § 8 hatten wir zwei Theorien praktischer Rationalität unterschieden: T1 und T2. Die ideale rationale Person bezüglich T1 verfügt über eine (subjektive) Bewertungsfunktion u („Nutzen") über alle mögliche Weltzustände (aus X) und eine subjektive Wahrscheinlichkeitsfunktion p („Glauben") über alle Mengen von Weltzuständen (PotX). Ihr Handeln maximiert den Erwartungswert von u bezüglich p. Die ideale rationale Person bezüglich T2 ist dagegen dadurch definiert, daß ihre Präferenzrelation bestimmte, in § 8 aufgeführte Kohärenzaxiome erfüllt. T1 und T2 sind nicht äquivalent. T2 garantiert allerdings die Existenz einer – bis auf lineare Transformation eindeutigen – Bewertungsfunktion, die die Präferenzen der Person repräsentiert. Wir haben diejenige Theorie praktischer Rationalität, die verlangt, daß eine rationale Person ihre Handlungen jeweils absichtlich so wählt, daß sie den Erwartungswert des (subjektiven) Nutzens der Handlungsfolgen maximiert, als „Bayes'schen Konsequentialismus" bezeichnet.

Die unterdessen vorgetragenen Argumente haben deutlich gemacht, daß der Bayes'sche Konsequentialismus keine überzeugende Theorie praktischer Rationalität darstellt. Auch moralische Qualifikationen der Folgenbewertung konnten diesen Befund nicht in Frage stellen. Ungeklärt sind jedoch die Implikationen, die dieses Ergebnis für die Kohärenzaxiome von T2 hat. Da man T2 als den Nucleus einer Kohärenztheorie praktischer Rationalität ansehen kann, steht damit nicht nur der Konsequentialismus, sondern auch die auf F. P. Ramsey zurückgehende Kohärenztheorie praktischer Rationalität zur Disposition. T2 kann nicht zur Rechtfertigung von T1 herangezogen werden, da T2 T1 nicht impliziert. Dennoch bleibt die beunruhigende Frage: Welche Rolle spielt die durch T2 konstituierte Bewertungsfunktion? Was optimiert der bezüglich T2 rationale Akteur, wenn er diese Funktion maximiert? In welchem Zusammenhang steht diese Funktion mit den subjektiven Präferenzen der handelnden Person? Aufgabe dieses Abschnittes ist es, auf diese Frage eine Antwort zu geben.

Man stelle sich zunächst einen idealen Kantischen Akteur vor, der seine Handlungspräferenzen – zumindest soweit es sich nicht um pragmatische, sondern um moralische Fragen handelt – an der Erfüllung des moralischen Gesetzes (des kategorischen Imperativs) orientiert. Man nehme wei-

ter an, daß es sich um einen „vernünftigen" Kantianer handelt, der zumindest in Fällen, in denen die Befolgung des kategorischen Imperativs katastrophale Folgen hätte, eine Handlung wählt, die nicht im Einklang mit dem kategorischen Imperativ ist. Für das Folgende ist die Annahme hinreichend, daß der kategorische Imperativ bisweilen Handlungen empfiehlt, deren Folgen nicht erwartungswertmaximierend bezüglich der subjektiven Zustandsbewertungen der handelnden Person sind. Der Kantische Akteur wird sich bisweilen für Handlungen entscheiden, deren Folgen (auch bei Berücksichtigung moralischer Wertungen) nicht optimal sind (bzw. – im probabilistischen Fall – deren Erwartungswert nicht maximal ist). Können die Präferenzen des so charakterisierten Kantischen Akteurs die Bedingungen der Kohärenztheorie praktischer Rationalität (T2) erfüllen?

Die Antwort lautet: *Wenn X die Menge der Weltzustände ist und die Präferenzen über Handlungen identifiziert werden mit Präferenzen über die den Handlungen zuzuordnenden Wahrscheinlichkeitsverteilungen über X, dann ist das Entscheidungsverhalten des Kantischen Akteurs i. d. R. nicht im Einklang mit den Kohärenzbedingungen von T2.*

Dies kann man sich leicht klarmachen: Eine Handlung h und eine Handlung h′ können die gleiche Wahrscheinlichkeitsverteilung über Weltzustände mit sich führen, und dennoch mag h durch den kategorischen Imperativ erlaubt und h′ verboten sein. Die Einordnung einer Handlung in die Präferenzrelation des Kantischen Akteurs erfolgt eben nicht nur aufgrund der durch die Handlungen induzierten Wahrscheinlichkeitsverteilungen über Folgen (bzw. über Weltzustände), sondern eben auch aufgrund von Konformität bzw. Defektivität gegenüber einem Kriterium (dem kategorischen Imperativ). Damit unvereinbar sind jedoch das Substitutions-, das Stetigkeits- und das Monotonieaxiom – also alle wesentlichen Kohärenzbedingungen von T2 bei probabilistisch erweiterter Alternativenmenge. Dieses Resultat gilt allerdings nur unter der Voraussetzung „wenn X die Menge der Weltzustände ist und die Präferenzen über Handlungen identifiziert werden mit Präferenzen über die den Handlungen zuzuordnenden Wahrscheinlichkeitsverteilungen über X", die man kurz als die Annahme der *Repräsentatierbarkeit von Handlungen durch Wahrscheinlichkeitsverteilungen über Weltzustände* bezeichnen kann.

Die Inkompatibilität von Kohärenztheorie und Kantischer Rationalität ist jedoch behoben, wenn man diese Art der Koppelung von Handlungen und Weltzuständen aufgibt. Da für die Kohärenztheorie praktischer Rationalität die Kohärenz von *Handlungs*präferenzen ohnehin grundlegend ist, (die die simultane Zuordnung von Glaubens- und Wünschbarkeitsgraden erlaubt), ist es nur folgerichtig, die Kohärenz der Handlungspräferenzen des Kantischen Akteurs zu prüfen, ohne diese Kohärenzprüfung mit der problematischen Koppelungsthese „Handlungen werden durch Wahrscheinlichkeitsverteilungen über Weltzustände repräsentiert" zu belasten.

Die in eine Präferenzrelation einzuordnenden Alternativen sind jetzt Handlungsoptionen. Zu prüfen ist also, ob *die Präferenzen des Kantischen Akteurs über Handlungsoptionen die Bedingungen der Kohärenztheorie praktischer Rationalität* erfüllen können.

X sei die Menge der Handlungsoptionen und X* sei die Menge der Aussichten über solche Optionen. Aussichten über X sind so zu interpretieren, daß – je nach Umstand – mit einer Handlung h eine Handlung h′ oder eine Handlung h″ etc. vollzogen wird. Z. B.: Ich betätige die Pumpe. Je nach Umstand kann diese Handlung die Versorgung der Anwohner mit Trinkwasser oder ihre Vergiftung bedeuten. Im zweiten Fall hätte ich mit der Betätigung der Pumpe zugleich eine andere Handlung, nämlich die des Vergiftens, vollzogen. Abschlußhandlungen im Sinne von § 2 können keine Aussichten über Handlungen zugeordnet werden, sie müssen also direkt in die Präferenzordnung der Person eingereiht werden. Dies wäre das Analogon zu Handlungen im Sicherheitsfall mit einem Weltzustand als sicherer Konsequenz.

Können die Präferenzen des Kantischen Akteurs über eine so interpretierte Alternativenmenge X* kohärent im Sinne von T2 sein? Für technische und pragmatische Imperative im Sinne Kants gilt dies ohnehin, aber auch bei der Abwägung, welche Handlung bei moralischer Bewertung zu vollziehen sei, geraten die Kohärenzbedingungen, angewandt auf Präferenzen über Handlungsoptionen, nicht notwendigerweise in Konflikt mit dem Gebot, nur nach derjenigen Maxime zu handeln, durch die man zugleich wollen kann, daß sie ein allgemeines Gesetz werde[8].

Es ist anzunehmen, daß auch der Kantische Akteur zwischen zwei Aussichten indifferent ist, wenn sich eine Aussicht in die andere durch mathematische Umformung nach dem Wahrscheinlichkeitskalkül überführen läßt, die Reduktionsbedingung ist also erfüllt. Weiterhin ist anzunehmen, daß auch der Kantische Akteur, wenn er zwischen einem x aus X und einer Aussicht a indifferent ist, überall x durch a ersetzen kann, ohne daß sich an ihren jeweiligen Präferenzen etwas ändert, die Substituierbarkeitsbedingung ist also ebenfalls erfüllt. Wenn der Kantische Akteur mit zwei Aussichten konfrontiert wird, deren Wahrscheinlichkeiten sich auf ein gleiches Paar von Handlungen aus X beziehen, dann wird er diejenige Aussicht vorziehen, die für die präferierte Handlung eine höhere Wahrscheinlichkeit vorsieht: $\forall x,y \in X : [x > y \rightarrow (p \geq p' \rightarrow \langle p/x \,\&\, (1-p)/y\rangle \langle p'/x \,\&\, (1-p')/y\rangle)]$, die Monotoniebedingung ist also ebenfalls erfüllt.

Probleme gibt es jedoch bei der Stetigkeitsbedingung. Zum einen, weil die Rede vom „subjektiv besten Weltzustand" schon eine Herausforderung an die Phantasie darstellt, aber die Rede von der „subjektiv besten Handlung" die Grenze zur Unverständlichkeit generell, insbesondere aber im Rahmen einer nicht-axiologischen Pflichtenethik wohl schon über-

[8] s. I. Kant, *Grundlegung zur Metaphysik der Sitten*. AA, S. 421.

schritten hat. Zum anderen, weil die Einordnung moralisch zu beurteilender Handlungen in die Präferenzrelation Schwierigkeiten aufwirft.

Betrachten wir zunächst den Fall einer bloßen moralischen Dichotomie: Es gibt moralisch erlaubte und moralisch verbotene Handlungen, je nachdem, ob die Maximen mit dem Kriterium des Kategorischen Imperativs übereinstimmen. Es gibt keine moralische Rangordnung unter Handlungen. Das moralische Gebot rangiert über allen anderen Handlungsbewertungen pragmatischer oder technischer Art. Die moralisch gebotenen Handlungen sind strikt vorgeordnet, d. h., wenn zwei Handlungen offenstehen und die eine moralisch geboten ist, dann ist diese zu wählen, wie positiv auch immer die Folgen der anderen Handlungen seien. Die Vorordnung der moralisch gebotenen Handlungen allein verletzt noch nicht die Stetigkeitsbedingung. Irritierend ist zunächst nur, daß alle moralisch gebotenen Handlungen nicht nur vorgeordnet sind, sondern daß zwischen diesen auch Indifferenz besteht. Die Stetigkeitsbedingung wird jedoch bei einer kategorischen Interpretation der Vorordnung verletzt: Angenommen, h_1 ist eine moralisch gebotene Handlung (bzw. $\neg h_1$ ist moralisch verboten, $\neg h_1$ ist nicht pflichtgemäß), während h_2 und h_3 moralisch erlaubt sind und von der handelnden Person unter pragmatischen Gesichtspunkten h_2 gegenüber h_3 vorgezogen wird. Diese Konstellation ist nur möglich, wenn mit h_2 und h_3 jeweils zugleich h_1 vollzogen wird, insofern findet nur eine Abwägung zwischen h_2 und h_3 statt (wäre der Vollzug von h_2 bzw. h_3 nicht mit dem Vollzug von h_1 verbunden, geschähe etwas moralisch Unzulässiges, nämlich $\neg h_1$, im Widerspruch zur Annahme). Die beste Handlung ist in jedem Falle eine moralisch gebotene, die schlechteste eine moralisch verbotene. Moralisch erlaubte Handlungen implizieren moralisch gebotene Handlungen. Es kann daher keine Aussichten geben, deren gutes Ergebnis eine moralisch gebotene und deren schlechtes Ergebnis eine moralisch verbotene Handlung ist, die aber indifferent zu einer moralisch erlaubten Handlung ist.

Die Vorstellung eines Kontinuums der Handlungsbewertung, deren oberen Teil moralisch gebotene, deren unteren Teil moralisch verbotene und deren Mittelteil moralisch erlaubte Handlungen bilden, ist bei einer kategorischen Variante Kantischer Ethik nicht aufrechtzuerhalten. Die Stetigkeitsbedingung ist dennoch formal nicht verletzt, da unter einer moralischen Beurteilung die Prämisse der Bedingung $x_a > x > x_z$ nicht erfüllbar und damit die Bedingung (ex falso quod libet) trivial erfüllt ist; dennoch ist zu beachten, daß es keine Wahrscheinlichkeit p im offenen Intervall (0,1) gibt, so daß $p/x \,\&\, (1-p)/z = y$, wobei x eine moralisch gebotene, z eine moralisch verbotene und y eine moralisch erlaubte Handlung ist. Eine außermoralische (pragmatische) Beurteilung, die die moralische Zulässigkeit immer schon voraussetzt, erlaubt ebenfalls keine Integration des moralischen und des außermoralischen Gesichtspunktes – innerhalb der außermoralischen Betrachtung gibt es allerdings keinen Grund anzunehmen, daß die Stetigkeitsbedingung nicht erfüllt ist.

Erst eine „konsequentialistische Abschwächung" kantischer Moral, die eine gewisse Abwägung zwischen Regeleinhaltung und Handlungsfolgen zuläßt und z. B. – entgegen Kant – zu lügen gestattet, wenn damit das Leben eines Unschuldigen gerettet werden kann, würde die Anwendbarkeit der Stetigkeitsbedingung und damit eine einheitliche numerische Repräsentation der Handlungspräferenzen ermöglichen. Die Verletzung der Bedingung der Verallgemeinerbarkeit einer Maxime könnte dann durch sehr gute Handlungsfolgen bzw. die Vermeidung einer Katastrophe aufgewogen werden. Im Sinne supererogativer Handlungen gäbe es auch innerhalb des moralischen Bereichs Abstufungen: Nicht nur die Konformität mit dem moralischen Gesetz, sondern auch etwa der Beitrag zum menschlichen Wohlergehen würden das moralisch Wünschenswerte bestimmen. Wir fassen diese Modifikation mit dem Begriff der *konsequentialistischen Abschwächung* zusammen. Handlungen, deren Maximen man als ein allgemeines Gesetz wollen kann, ließen sich dann in die Präferenzrelation kohärent einordnen, sie wären qua konsequentialistischer Abschwächung durch eine simultane Zuordnung entsprechender Glaubens- und Wünschbarkeitsfunktionen repräsentierbar.

Der Übergang vom Kohärenzmodell zum konsequentialistischen Optimierungsmodell bleibt jedoch auch bei einer konsequentialistisch abgeschwächten Kantischen Ethik ausgeschlossen. Die beiden entscheidenden Schritte dieses Übergangs sind nicht realisierbar: *Erstens lassen sich den Handlungen nicht Wahrscheinlichkeitsverteilungen über Weltzustände im üblichen Sinne zuordnen, und zweitens läßt sich u als quantitative Darstellung der Handlungspräferenzen nicht als Repräsentation der Handlungsmotive interpretieren. Die Zuordnung quantitativer Werte zu den Elementen aus X* repräsentiert nur die kohärente Präferenzrelation über Handlungen und repräsentiert nicht die subjektive Wünschbarkeit von Weltzuständen im üblichen Sinne.*

Ich habe im letzten Absatz von „Weltzuständen im üblichen Sinne" gesprochen, da es natürlich möglich ist, den Handlungen „formal" Wahrscheinlichkeitsverteilungen über Weltzustände zuzuordnen, sofern die Beschreibungen dieser Weltzustände Merkmale enthalten wie „die Handlung der Person erfüllte den kategorischen Imperativ". Mit dieser Modifikation könnte man dann allerdings die durch das Kohärenzmodell konstituierte Bewertungsfunktion nicht mehr in das Optimierungsmodell im konsequentialistischen Sinne überführen, denn dann sind es nicht Weltzustände, die optimiert werden, vielmehr wird bestimmten Merkmalen der Handlung selbst ein intrinsischer Wert zugeordnet.

Aus Achtung für das moralische Gesetz zu handeln, stellt einen möglichen Handlungsgrund dar. Wie oben gezeigt, ist dieser Handlungsgrund unter bestimmten Bedingungen (konsequentialistische Abschwächung) mit *Ramsey-Kohärenz* – wie wir es i. F. nennen wollen – vereinbar: Die Präferenzen einer Person können Ramsey-kohärent sein, obwohl sie nicht

durch die Zielsetzung der Folgenoptimierung, sondern etwa durch den Handlungsgrund „moralisches Gesetz" motiviert sind.

Da das Kooperationsproblem im Zentrum unserer Konsequentialismus-Kritik in den Kapiteln 11 bis 14 stand, soll abschließend anhand des paradigmatischen Falles des Gefangenendilemmas das Verhältnis von Folgenoptimierung und Kohärenz bei kooperativem Handeln dargestellt werden.

Schon in den 60er Jahren wurden empirische Untersuchungen darüber angestellt, wie sich Personen in Interaktionssituationen vom Typ des Gefangenendilemmas verhalten[9]. Dabei stellte sich heraus, daß Personen auch dann, wenn sie über die Struktur der Interaktionssituation informiert sind und wissen, daß es sich um eine einmalige Interaktion handelt („One-shot-PD-game"), sich zu einem erstaunlich hohen Prozentsatz kooperativ, also nicht individuell optimierend, verhalten. Wenn es auch nicht die einzig mögliche ist, liegt doch folgende Interpretation eines kooperativen Verhaltens in Interaktionssituationen vom Typ des Gefangenendilemmas nahe: Die Person kooperiert, weil sie weiß, daß beidseitig kooperatives Verhalten kollektiv rational ist. Nur unter sehr speziellen Annahmen ließe sich die Vermutung anstellen, daß altruistische Motive eine Rolle spielen. (Durch entsprechende Variationen der Auszahlungen ließe sich sogar empirisch eruieren, ob altruistische Motive dabei eine Rolle spielen oder nicht.) Zumindest scheint es intuitiv in keiner Weise abwegig zu sein anzunehmen, daß eine Person sich bewußt für die kooperative Strategie entscheidet, obwohl ihre subjektiven Bewertungen der vier möglichen Ergebnisse einer vorgegebenen Interaktionsstruktur dem Typ des Gefangenendilemmas entsprechen.

Um eine petitio principii auszuschließen, darf das Gefangenendilemma nicht über die Rangordnung der Entscheidungskombinationen, sondern ausschließlich über die Rangordnung der Konsequenzen der Entscheidungskombinationen definiert werden. In der ursprünglichen Geschichte, die zum Namen „Gefangenendilemma" geführt hat, waren die Konsequenzen Gefängnisjahre für die beiden Interaktionsbeteiligten. Derjenige, der nicht gestand, erhielt etwa zwei Jahre Gefängnis, wenn der andere ebenfalls nicht gestand, während er als Nicht-Geständiger besonders hart bestraft wurde, wenn der andere gestand und damit beide ihres Verbrechens überführt werden konnten, etwa 12 Jahre. Derjenige, der gestand, wurde als Kronzeuge der Anklage freigesprochen, wenn der andere nicht gestand, und erhielt 10 Jahre, wenn der andere ebenfalls gestand. „Gestehen" war hier die nicht-kooperative Strategie, im folgenden mit „n" abgekürzt, und „nicht gestehen" war die kooperative Strategie, im folgenden mit „k" abgekürzt. Die Konsequenzenmatrix (Konsequenzen in Gefängnisjahren) sieht also folgermaßen aus:

[9] Vgl. A. Rapoport/M. J. Guyer/D. G. Gordon, *The 2x2 Game*, Ann Arbor 1976, sowie neuere Erhebungen in L. Kern/H.-G. Räder, Private Goods and Public Bads, in *Theory and Decision – in Honour of W. Leinfellner*, hg. von G. L. Eberlein u. Berghel, Dordrecht 1988.

		B	
		n	k
A	n	10/10	0/12
	k	12/0	2/2

Wir sagten, die Struktur des Gefangenendilemmas definiere sich über die je subjektiven Präferenzen über Konsequenzen. Insofern handelt es sich bei dieser Interaktionsstruktur mit den gegebenen Konsequenzen (in Gefängnisjahren) nur dann um eine Interaktionsstruktur vom Typ des Gefangenendilemmas, wenn A es auch dann vorzieht freizukommen, wenn B dadurch 10 Jahre länger sitzen muß, und wenn A es auch dann vorzieht, seine Gefängniszeit um zwei Jahre zu verkürzen (von 12 auf 10), wenn dadurch die Gefängniszeit von B von 0 auf 10 Jahre zunimmt. Mit C_A für die A betreffenden Konsequenzen und C_B für die B betreffenden Konsequenzen ergibt sich also folgende, eine Interaktionsstruktur vom Typ des Gefangenendilemmas definierende Präferenzrelation:

A: $C_A(n,k) \;/\; C_B(n,k)$
$C_A(k,k) \;/\; C_B(k,k)$
$C_A(n,n) \;/\; C_B(n,n)$
$C_A(k,n) \;/\; C_B(k,n)$

B: $C_A(k,n) \;/\; C_B(k,n)$
$C_A(k,k) \;/\; C_B(k,k)$
$C_A(n,n) \;/\; C_B(n,n)$
$C_A(n,k) \;/\; C_B(n,k)$

Diese umständliche Notation wurde bewußt gewählt, um zu vermeiden, daß die Wahl der symbolischen Darstellung eine theoretische Vorentscheidung für das *revealed preference*-Konzept, also die Annahme, daß sich die Präferenzen im Entscheidungsverhalten offenbaren, beinhaltet. Wenn wir uns bei der Darstellung der Präferenzen bei A auf die linke und bei B auf die rechte Spalte, also auf die Konsequenzen für A bzw. für B, beschränkt hätten, dann wäre damit schon implizit die Annahme ausschließlich eigenorientierter Präferenzen eingeflossen.

A: $C_A(n,k)$
$C_A(k,k)$
$C_A(n,n)$
$C_A(k,n)$

B: $C_B(n,k)$
$C_B(k,k)$
$C_B(n,n)$
$C_B(k,n)$

Wenn wir die Rangfolge der Strategienkombinationen zur Charakterisierung einer Interaktionssituation vom Typ des Gefangenendilemmas herangezogen hätten, dann wäre damit eine Vorentscheidung für die individuell-optimierende Rationalitätskonzeption getroffen worden.

A: n,k
k,k
n,n
k,n

B: k,n
k,k
n,n
n,k

Wir betrachten nun erneut die Person, die in Kenntnis der Interaktionsstruktur, in der sie sich befindet, und in Kenntnis der C-Präferenzen (der Präferenzen, die sich auf die Konsequenzen der Interaktion beziehen), sich für die kooperative Strategie entscheidet. Als Handlungsgrund (als Rechtfertigung dieser Entscheidung) gelte dieser Person die Tatsache, daß beidseitige Kooperation beidseitiger Nicht-Kooperation vorzuziehen ist, was vor dem Hintergrund einer symmetrischen Interaktionssituation für die Wahl der kooperativen Entscheidung spreche. Die Präferenzen einer solchen Person lassen sich kohärent offensichtlich nicht durch die zuletzt angegebene Präferenzstruktur wiedergeben. Denn wenn wir annehmen, daß alle wesentlichen Informationen der vier Alternativen durch die jeweilige Strategienkombination gegeben sind, dann hieße dies, daß etwa A sich bei der Alternative (n,k) und (k,k) für (k,k) entscheidet, was der angegebenen Präferenzordnung zuwiderläuft. Die Präferenzen einer solchen Person wären also in der Tat inkohärent.

Die Absicht zu kooperieren kann einen guten Handlungsgrund darstellen. Es ist jedoch in der Regel adäquat, kooperative Handlungsgründe als Umwertung von Konsequenzen zu interpretieren[10]. Ein rationaler Akteur kann sich auch dann für eine kooperative Strategie in Interaktionssituationen vom Typus des Gefangenendilemmas entscheiden, wenn er die Konsequenzen, die Nicht-Kooperation (bei gegebener Entscheidung des anderen Interaktionsbeteiligten) hat, jeweils vorzieht. In diesem Beispiel heißt das: Es ist möglich, daß sich eine rationale Person für die kooperative Strategie entscheidet (nicht gestehen), auch wenn sie das Ergebnis 0 Gefängnisjahre für sich selbst und 12 Gefängnisjahre für den anderen dem Ergebnis 2 Gefängnisjahre für sich selbst und 2 Gefängnisjahre für den

[10] Diese Interpretation entspräche dem Ansatz der Metapräferenzen, wie sie A. K. Sen, Choice, Ordering and Morality, in: *Practical Reason*, hg. von S. Körner, Oxford 1974, S. 54–67 entwickelt hat. Ich kritisiere diesen Ansatz eingehender in „Practical Reason or Metapreferences? An Undogmatic Defence of Kantian Morality", *Theory and Decision* 30 (1991), S. 133–162.

anderen ebenso vorzieht wie 10 Gefängnisjahre für sich selbst und 10 Gefängnisjahre für den anderen dem Ergebnis 12 Gefängnisjahre für sich selbst und 0 Gefängnisjahre für den anderen. Ja, noch mehr: Eine Person kann aus logischen Gründen kooperative Handlungsabsichten nur haben, wenn sie keine Umwertung der Konsequenzen vornimmt, die die Gefangenendilemma-Struktur der Interaktionssituation aufhebt. *Konsequentialistische Optimierung und Kooperation sind nicht kompatibel.* Kooperative Handlungsgründe erzwingen andererseits keine inkohärenten Präferenzrelationen: *Ramsey-Kohärenz und Kooperation sind kompatibel.*

§ 52 Lebensform und Gesellschaftsform

Im Verlauf unserer Kritik des Konsequentialismus wurde die ursprüngliche heuristische Trennung einer ethischen und einer rationalitätstheoretischen Betrachtungsweise zunehmend aufgehoben. Dies zeigte sich besonders deutlich beim Übergang zu einer intrapersonellen Betrachtung: Die strukturspreng ende Wirkung konsequentialistischer Optimierung ist nicht nur ein Problem der gesellschaftlichen Kooperation, sondern auch der individuellen Lebensgestaltung. Während die Philosophie der griechischen und römischen Klassik die Frage der richtigen Lebensgestaltung als das zentrale ethische Problem behandelte, ist das philosophische Interesse an dieser Frage in der neuzeitlichen und modernen Philosophie gering. Manche Defizite der jüngeren handlungsorientierten Ansätze philosophischer Ethik haben wohl darin ihre Ursache. Ein deutliches Beispiel dafür sind die konsequentialistischen Ansätze in der Ethik.

Es ist legitim, eine ethische Theorie auch daran zu messen, welche Eigenschaften eine Gesellschaft von Personen hätte, die dieser Theorie in idealer Weise genügen. In der idealen konsequentialistischen Gesellschaft steht die Gestaltung des eigenen Lebens sowohl äußeren wie inneren Ansprüchen gegenüber: Die *äußeren Ansprüche* ergeben sich aus der konsequentialistischen Forderung, alle abschätzbaren Handlungsfolgen zu berücksichtigen, und das schließt die Wirkung eigenen Handelns für das Handeln anderer Personen ein. Diese äußeren Ansprüche machen eine vernünftige Lebensplanung unter normalen Umständen unmöglich, da sie die Verfolgung persönlicher Ziele und Projekte sowie die Aufrechterhaltung persönlicher Bindungen unterminieren. Wir haben dieses Phänomen unter dem Aspekt der Integrität der Person, aber auch im Zusammenhang mit individuellen moralischen Rechten diskutiert. Die *inneren Ansprüche* konsequentialistisch rationalen Handelns sprengen in Gestalt punktueller Optimierung die Strukturen der eigenen Lebensgestaltung (s. Kap. 10).

Wenn man die Ergebnisse der Kapitel 11, 12 und 13 zusammenfaßt, dann kann man das Zentrum unserer Kritik als das *doppelte* (interpersonelle und intrapersonelle) *Koordinationsproblem einer konsequentialistischen Konzeption praktischer Rationalität* bezeichnen. Koordination ergibt

sich aus der sinnvollen Einfügung individuellen (punktuellen) Handelns in bestimmte Handlungsstrukturen. Konsequentialistisch rationales Handeln führt jedoch nicht zu wünschenswerten Strukturen individuellen und gesellschaftlichen Verhaltens.

Die beiden Orientierungspunkte des ethischen Diskurses, der eines *guten Lebens* und der einer *wohlgeordneten Gesellschaft*, sind in einer komplexen Weise miteinander verknüpft, der der konsequentialistische Ansatz nicht gerecht werden kann[11]. Die konsequentialistische Theorie praktischer Rationalität deckt zwar, je nach Wertfunktion, das Spektrum von der eigenorientierten Optimierung bis zur individuellen Selbstaufopferung für das gesellschaftliche Gute ab. Die Untersuchung hat jedoch gezeigt, daß diese Theorie in keinem Bereich dieses Spektrums überzeugend ist. Der Konsequentialismus bewältigt beide Aufgabenstellungen einer Theorie praktischer Rationalität nur unzureichend: Weder bietet er taugliche Kriterien für die Gestaltung eines guten Lebens noch solche für die sinnvolle Koordination gesellschaftlichen Handelns und individueller Lebensformen.

Die naheliegende Reaktion auf die Koordinationsproblematik konsequentialistischer Rationalität besteht in der Beschränkung der Handlungsoptionen. Wer weniger zu entscheiden hat, kann in vielen Fällen wenigstens nicht falsch entscheiden. Das ist die Quintessenz sowohl der unterschiedlichen Praktiken der Selbstbindung wie der Etablierung von Sanktionen in foro externo und foro interno. Im Rahmen des Konsequentialismus wird ein Mehr an Rationalität mit einem Weniger an Entscheidungsfreiheit erkauft. Es spricht gegen eine Rationalitätskonzeption, wenn rationale Personen wünschen müssen, nicht (rational) entscheiden zu können.

Die Lebensform ist weitgehend von den Strukturen individuellen und punktuellen Handelns geprägt. Diese in vielen Fällen selbst gewählten oder zumindest beeinflußbaren Strukturen bestimmen die Qualität der individuellen Lebensform ebenso wie die der Gesellschaftsform als Ganzes, die man sich als eine Vernetzung individueller Lebensformen durch unterschiedliche Formen der Interaktion denken kann.

Ein mögliches Handlungsmotiv ist das der *Koordination:* sich je individuell und punktuell so zu verhalten, daß die Einzelhandlung sich in eine befürwortete Struktur individuellen bzw. gesellschaftlichen Handelns einfügt. Der Gefangene aus § 45 kann das Motiv haben, sich so zu verhalten, wie er sich wünscht, daß sich auch der andere Gefangene verhält. Das *Motiv der Koordination* kann ein guter Grund für eine Entscheidung sein. Eine adäquate Theorie der Handlungsbegründung sollte diese spezifische

[11] Die klassische Theorie der Übereinstimmung von Gerechtigkeit und gutem Leben, wie sie z. B. in Platons *Gorgias*-Dialog skizziert ist, ist sicher zu optimistisch, aber sie thematisiert dieses ethische Problem, während das moderne Schisma von gutem Leben und moralischer Pflicht eine Umgehung des Problems darstellt.

Art *normativischer Gründe* nicht ausschließen. So gibt es das Phänomen der koordinativen Absichten, die sich in bestimmten Interaktionssituationen ausprägen und sich etwa in folgender Weise rekonstruieren lassen: Eine Person aus einer Gruppe hat eine *koordinative Absicht*, wenn sie ihren Teil zu einer kollektiven Handlung beitragen will, sofern sie annehmen kann, daß jedes andere wohlinformierte und rationale Mitglied der Gruppe seinen Teil ebenso beiträgt[12]. Koordinative Absichten können sich aber auch auf die Gestaltung des eigenen Lebens beziehen: Ich verhalte mich punktuell so, wie es der von mir am meisten geschätzten Form des Lebens entspricht.

Zwischen den unterschiedlichen Strukturen des Handelns gibt es wiederum eine strukturelle Beziehung: Die einzelnen Strukturen fügen sich zu Strukturen größerer Allgemeinheit – individuelle Lebensformen vernetzen sich zu einer Gesellschaftsform. Wenn man die einzelne Person nicht als Konglomerat von Augenblicksempfindungen, Wahrnehmungen und Reaktionen begreift[13], dann ist sie Träger der intrapersonellen Strukturierung ihres (je punktuellen) Handelns und der Koordination dieser individuellen Lebensform mit den Lebensformen anderer Personen. Die moralische Person befindet sich, bildlich gesprochen, an der Nahtstelle zweier unterschiedlicher Ebenen der Handlungskoordination.

Die Strukturen des Handelns sind weder intra- noch interpersonell (durch Erziehung, Sanktionen, Konventionen, anthropologische Determinanten etc.) in toto vorgegeben, sondern zu einem guten Teil Gegenstand freier Entscheidung – sie können geformt, entwickelt, verändert oder bewahrt werden. Es stellt sich daher eine doppelte Aufgabe: Kriterien für ethisch wünschbare Strukturen[14] und Kriterien für die Einfügung der individuellen Handlung in diese Strukturen zu entwickeln[15] – die konsequentialistische Theorie praktischer Rationalität und die auf ihr beruhenden konsequentialistischen Ethiken werden dieser doppelten Aufgabe nicht gerecht.

[12] Näheres dazu in R. Tuomela, *A Theory of Social Action*, Kap. 4 und Verf. „Wir-Intentionen und die Logik kollektiver Entscheidungen", *Analyse und Kritik* 8 (1986), S. 96–108.

[13] Vgl. die Diskussionsbeiträge in: *The Multiple Self*, hg. von J. Elster, Cambridge 1986.

[14] Das umfaßt unter einem interpersonellen Aspekt u. a. die axiologische Theorie gesellschaftlicher Institutionen und unter einem intrapersonellen Aspekt die Theorie der guten Lebensform.

[15] Unter der Fragestellung, welches Verhältnis zwischen Selbstliebe und Vernunft bzw. Gerechtigkeit besteht, ist im Lauf von zweieinhalb Jahrtausenden vieles gedacht und geschrieben worden, was die zeitgenössische Debatte um praktische Rationalität bereichern könnte. Als zwei Beispiele für viele vgl. auf der einen Seite den zweiten Brief von Alexander Pope, *An Essay on Man. Being the First Book of Ethic Epistels* (London 1733), wo Pope davon ausgeht, der Mensch habe ein natürliches Interesse am Wohl des anderen, das menschliche Geschlecht sei durch eine „Kette der Liebe" zu einem harmonischen Ganzen verbunden; und auf der anderen Seite Duc de La Rochefoucauld, der auch das scheinbar selbstlose Verhalten als in Wirklichkeit irrational-egoistisches, unfreies Handeln geistreich analysiert, s. *Reflexions ou sentences et maximes morales*, Paris 1665, dt. *Gemüths-Spiegel* ... (1669).

Register

Namenregister

Adams, R. M. 86
Aldrich, J. 103
Anscombe, A. M. 24, 32
Aqvist, L. 12, 20
Aristoteles 35, 42, 55, 89, 152
Armstrong, W. E. 126
Arrow, K. J. 1, 67 ff.
Audi, R. 49
Austen-Smith, D. 103
Austin, J. L. 25, 28
Axelrod, R. 157

Baier, A. 15
Barnes, J. 103
Beckermann, A. 32
Bennett, J. 25, 27, 32
Bentham, J. 56, 75
Birnbacher, D. 59, 113
Black, M. 159
Brand, M. 15
Brandt, R. B. 116
Brentano 24
Brink, D. 89
Broad, C. D. 47
Broderick, J. 140
Brody, B. 117
Buchanan, J. 97
Butler, J. 47, 55, 79

Cheney, D. 47
Chisholm, R. M. 3, 15, 24, 32

Daniels, N. 60, 61
Danto, A. C. 15
Davidson, D. 13, 68
Dostojewski, F. 169
Downs, A. 46

Edgeworth, F. Y. 1
Elster, J. 169, 188
Epikur 58
Ewing, A. C. 116

Feinberg, J. 30
Feldmann, A. 136
Feyerabend, P. 140
Foot, P. 32
Frankena, W. K. 63
Frege, G. 24

Gauthier, D. 97, 157
Gean, W. 32
Georgescu-Roegen, N. 126
Gibbard, A. 68, 103, 135
Gody, A. 13
Goldmann, L. 13
Goodman, L. A. 126
Green, J. 136
Grice, H. P. 25, 58
Guyer, J. 183

Hammond, P. F. 61
Hanh, F. H. 1
Hare, R. M. 55, 113, 159, 174
Harrison, J. 116
Harrod, R. F. 55, 56, 57
Harsanyi, J. C. 89
Hart, H. L. A. 30, 95, 96
Heraklit 58
Hobbes, Th. 143, 155, 157, 165
Hodgons, D. H. 54, 139, 146
Hoerster, N. 116, 119, 159
Höffe, O. 95, 148
Hume, D. 79, 159
Hutcheson, F. 79

Jordan, P. 22

Kant, I. 22, 55, 83, 85, 91, 97, 143, 144, 177 ff.
Kelly, J. 103
Kelsey, D. 103, 104
Kenny, A. 32
Kern, L. 40, 70, 103, 104, 109, 130, 183
Kirsch, G. 1
Kneale, W. 24
Kutschera, F. V. 20, 80

Laffont, J. J. 136
Lenk, H. 20
Lewis, D. 13, 25, 114
Locke, J. 143

Lomansky, L. 89
Luce, R. D. 1
Lyons, D. 116, 117

MacIntyre, A. 64, 104
Mackie, J. L. 55, 127, 131, 155, 157, 163
Markowitz, H. 126
Marquard, O. 148
Martin, J. R. 15
McManus, M. 61
Meggle, G. 13
Melden, A. I. 15, 22, 32, 61
Mele, A. 49
Mezzetti, C. 103
Mill, John St. 54, 56, 76
Moore, G. E. 55, 56, 79, 104

Narveson, J. 116
Nowell-Smith, B. P.-H. 80
Nozick, R. 64, 65, 97, 143

Olson, M. 1, 46

Parfit, D. 58, 59
Peacock, A. T. 103
Pitcher, G. 30
Platon 55, 89, 152, 187
Plott, C. 136
Pope, A. 188
Prior, A. N. 24
Putnam, H. 20

Räder, H.-G. 183
Raiffa, H. 1
Ramsey, F. P. 178, 182, 186
Rawls, J. 61, 62, 63, 80, 89, 97, 139, 143
Regan, D. 112, 116
Roberts, T. A. 47
Rochefoult, Duc de La 188
Rogerson, K. 90
Ross, W. D. 89, 116
Rothenberg, J. 126
Rowley, C. K. 103

Savage, L. J. 16, 17, 18, 19, 46f.
Savingy, E. v. 25, 26

Schaftesbury 79
Scheffler, S. 60, 77, 92
Schelling, T. C. 114
Schmeidle, D. 136
Searle, J. R. 24, 25, 27, 28
Sen, A. K. 61, 62, 68, 102, 103, 107, 135, 185
Sidgwick, H. 74
Simon, H. 37
Singer, M. G. 116
Smart, J. C. 54
Sneed, J. D. 19
Spaemann, R. 25, 76, 90, 148
Spinoza 81
Spohn, W. 12, 19, 20
Stegmüller, W. 1, 19, 20, 25
Stenlund, S. 20
Stocker, M. 89
Stout, A. K. 116
Stoutland, F. 15, 32
Sugden, R. 103
Suppes, P. 53, 67
Suzumura, K. 103

Taylor, C. 32
Taylor, R. 15, 32
Teichmüller, G. 142
Toulmin, S. 116
Trapp, R. 54, 62
Tuomela, R. 13, 58, 188

Voegelin, E. 142

Walter, J. 142
Walton, G. M. 61
Weber, M. 139
Wessels, U. 154
Wilcox, W. 89
Williams, B. 60, 89, 90, 93, 130, 139, 144
Wittgenstein, L. 21
Wright, G. H. v. 14, 20, 21, 26, 35, 175
Wriglesworth, J. L. 103
Wunderlich, D. 25

Sachregister

Abschlußhandlungen 15f., 33, 180ff.
Absichten 31ff., 42, 44, 47f., 51, 54, 85f., 148, 160, 168
- als Bestandteil von Handlungen 32
- Erfüllungsbedingungen von - 31ff., 42
- motivierende - 31ff., 42, 47f., 51ff., 148, 168
- rationale - 44
- verbunden mit Glauben 44
- verbunden mit Wünschen 44
- vorausgehende - 31ff., 42ff., 47f., 148, 160
- übergeordnete - 34
Aggregation 57f., 124ff., 132, 148
- Linearität der - 124, 127
- Linearitätseigenschaft 124
- nicht-lineare - 126
- inter- und intrapersonelle - 148
Aggregationsfunktion 57, 77
- regel 101, 107, 132f., 138
- regel, als kollektive Entscheidungsfunktion 105
- wert 77
Akrasia 138, 167, 170
Akteur 76, 83, 111ff., 122, 137, 140ff., 153f., 174f., 178ff.
- anarchistischer - 140
- kantischer - 178ff.
- konsequentialistisch rationaler - 137, 153f.
- moralischer - 76, 83, 11ff., 140ff., 174ff.
- rationaler - 154, 178
- realistischer - 140
Anthropologie 60, 151, 159, 165, 188
Äquivalenzklassen 21, 70
Arrows' Theorem 133, 134, 136
Arrowsche Alternative 69
- Zustände 68, 70, 71
Atomhandlungen 15ff.
Auswahlfunktionen, strikt kollektive 134ff.
axiologisch konsequentialistische Ethiken (s.a. Konsequentialismus) 86
axiologische Prädikate 81ff.
- Urteile 81ff.

Basishandlungen 15ff., 34
Bayes'sche Entscheidungstheorie 43, 81
- Lösung 121
- Subjektrelativierung 111
- r Konsequentialismus 36f., 39, 45, 48, 111
- s Modell 37, 45
Bayesianismus 36f., 39, 43, 45, 48f., 73f.
- ethischer - 49, 57, 73
- konsequentialistischer - 49
Begründungen 172ff.
- imperativische 176
- normativische 176
- objektive 172ff.
Bewertungsfunktion 40, 45, 67, 114ff., 121
- individuelle - 114
- interpersonell invariante - 114f.
- quantitative - 40
- subjektive - 45

Defektivität 179f.
demokratisches Entscheidungsverfahren 105f.
deontologische Ethik 76, 83, 86ff., 92
deontologische Intuition 96f.
- Theorien 63ff.
deontologischer Konsequentialismus 63ff., 110
diachronisches Vorgehen 128
Dispositionen 48ff., 130, 151, 159, 165, 188
Dominanzprinzip 168f.
Durchschnittsnutzenmaximierung 101, 115
dynamische Betrachtungen 127ff.

Effizienzkriterien 63
egoistische Wertfunktion 80
Einstellungen, propositionale 43ff.
Entscheidung, Kriterien der 47ff.
Entscheidungen 11, 17ff., 47ff., 68ff., 94, 97, 117, 128, 132ff., 136f., 140ff., 147, 154f., 162ff., 168ff.
- erster Ordnung 97
- individuelle - 132ff.
- irrationale - 162ff.
- kohärente - 53ff.
- kollektive - 68ff., 132f.
- konsequentialistisch rationale - 147ff.

- moralische – 53
- nicht kooperative – 163 ff.
- singuläre – 117
Entscheidungsfindung 17 ff., 94, 128, 136 f., 140 ff., 147, 154 f., 168
- und Ethik 140
- Regeln kollektiver – 147
- Theorie der richtigen – 141, 142
Entscheidungskoordination 128 ff.
- kriterium, konsequentialistisches 94, 137, 155, 168 f.
- oligarchie 137
- verfahren, öffentliches 132
Entscheidungsrationalität 37, 43 ff.
Entscheidungstheorie 36, 44 ff.
- als reduktionistische Theorie der Rationalität 44 ff.
Erfüllungsbedingungen (von Sprechakten) 29
Erwartungswertmaximierungskriterium 152
erweiterte Sympathie 67 ff.
Ethik 19, 66, 73 f., 76, 79 ff., 82 ff., 86 ff., 90 ff., 108, 113 ff., 130, 138 ff., 140 ff., 149, 151 f., 161 ff., 186 ff.
- deontologische – 76, 83, 86 ff., 92
- explanatorische – 79, 82
- Handlungs- und Motiv- 77, 170
- kantische – 66, 158, 181 f.
- klassische – 152 f.
- konsequentialistische – 53 ff., 73 f., 76, 79, 90, 96, 108, 130, 142, 186 ff.
- konstruktivistische – 144
- Motivations- 53 ff., 84
- naturalistische – 79 ff.
- normative – 64, 138 f., 151, 154, 172
- Pflichten – 83, 96
- praktische Vernunfts- 162 ff.
- psychologische – 82
- rigide unpersönliche – 90 f.
- soziologische – 79, 82
- subjektivistische 66, 81 ff., 107, 163 ff.
- teleologische – 77 f., 170 ff.
- Tugend- 84, 86 f., 111
- utilitaristische – 59 f., 63
Existentialismus 99
explikative Theorien 44
extrinsisch Gutes 42

Folgen 18 ff., 84 f., 91, 132 f.
- generelle Optimierung der 84
- Maximierung der 91
- punktuelle Optimierung der 132 f.
Freiheitsgrundsatz 64

Gefangendilemma 113, 127, 153 ff., 164, 183 ff.
Gerechtigkeit 23, 60 ff., 73 ff., 77, 11 ff., 117, 139, 143 ff.
- Fragen der – 60
- Prinzipien der – 61 ff.
- Theorie der – 64
- Verteilungs- 60 ff.
Gerechtigkeits-Empfinden 117
- konsequentialismus 66, 77, 117
- konzeption, kontraktualistisch begründetes – 143
- kriterien 61, 63
- prinzipien 139
- relation 73 f.
- sinn 139, 145
- theorie 73
- utilitarismus 117
Gesamtnutzen 62, 76, 115
- utilitaristisches Prinzip der -maximierung 62
Gesamtwohlmaximierung 23, 75
Gesellschaft 130, 138, 146, 186 ff.
- Instabilität der – 138
- konsequentialistische – 138 ff.
- Stabilität der – 139, 151
- wohlgeordnete – 187 f.
- Zustände in der – 62 ff.
Gesellschaftsform 186 ff.
Gibbard-Satterthwaite-Theorem 130

Handeln, normativ-ethische Theorie des -s 53 ff.
Handlung, Regelkonformität der 51
Handlung- und Motivethik 86
Handlungen 12 ff., 24 ff., 30 ff., 44 f., 50 ff., 74 ff., 84, 91 ff., 111 ff., 122 ff., 147, 152 ff., 160 ff., 171 ff.
- absichtliche 29 f., 33, 42, 168
- als intentionales Verhalten 29 ff.
- durch eine Handlung 12, 13, 14
- Erfüllungsbedingungen von – 33 ff., 48
- Erwartungswert von – 44
- externe Effekte von 100 ff.
- extrinsische – 48, 50
- identische 13
- individuelle – 76, 122, 54, 187 f.
- intrinsische – 42, 48 ff.
- kausale Wirkung von 74

- kollektive 100, 123, 125, 188
- komplexe – 34
- Konjunktion von 16
- Konsequenzen von – 32, 46, 49 ff.
- moralisch gebotene 181
- potentielle 16
- singuläre 16 f., 84
- supererogative 182
- unabsichtliche 29 f.
- unmoralische 160 f.
- Verantwortung für – 30 ff.
- wohlbegründete 177

Handlungsabsichten 24 ff., 186
- begleitende 24 f.
- kooperative 186
- motivierende 24 f.
- vorausgehende 24

Handlungsethik 84
- gründe 40 ff., 162, 171 ff., 181 f., 187
- konsequenzen 12, 13, 16, 18, 76, 168
- motivation 42, 84, 91 f., 183 ff.
- normen 163
- option 23, 76, 169, 170
- orientierung 92, 143 ff., 148, 170
- präferenzen 19, 152 f., 178 f., 182 f.
- utilitarismus 113, 116, 147
- theorie 15
- ziele 162, 171 ff.

Handlungsrationalität 36 ff., 43 f., 46 ff.
- konsequentialistische Interpretation 36 ff.

Handlungsziele 45, 55
hedonistische Wertfunktion 80
Hobbes-Variante 155 ff.

ideale moralische Person 62, 76, 83, 91, 111 ff., 141 ff., 154, 174 f.
ideale rationale Person 36 ff.
Identitätskriterien 58 ff.
individuelle Bewertungsfunktion 114
individuelle Rechte 95 ff., 151
Institutionen 96 f., 99, 111, 144, 146 ff., 177
- moralische – 144, 146 ff.
- normative – 177
- Regelungen durch – 96 f.

Integrität der Person 92 ff., 151, 186 f.
Intensionalität 14, 24 f., 29, 160 f.
intentionales Verhalten 29 ff.
Intentionen 31 ff., 45 f.
Interaktion 153 ff., 164, 170, 184
interpersonell invariante Bewertungen 114
interpersonell variante Bewertungen 67
intrinsisch Gutes 42
Intuitionen, moralische 59
Irrelevanzbedingung 104

Kantianismus 143, 156 ff., 177 ff.
kantische Ethik 66, 158, 181 f.
kantischer Konstruktivismus 143 ff., 177 ff.
kategorischer Imperativ 179, 181
Kohärenz 45 ff., 178 ff., 182 ff.
- axiom 178
- modell 182 ff.
- theorie praktischer Rationalität 178 ff.
- der Wünschbarkeit 45

kollektiv irrationale Strategien 164 ff.
kollektive Auswahlfunktionen 134 ff.
- Entscheidungen 107, 132 ff.
- Entscheidungsfindung 98, 99, 101, 105, 107
- Handlungen 100 f., 123, 125, 188
- Irrationalität 152 ff.
- Optimierung 121
- Präferenzrelationen 104 f., 107, 133
- Rationalität 156 ff.
- Wertfunktionen 165
- Wohlfahrtsfunktion 104, 107, 133

Kolmogorov-Axiome 81
Kommunikationsnormen 129 ff.
Konnexität, als Bedingung von Präferenzen 37 f.
Konsequentialismus 16, 36, 40, 46 ff., 50, 63 ff., 75 ff., 96, 110, 115 ff., 130 ff., 143, 145 ff., 164 ff.
- abgeschwächter – 66 ff., 77, 115
- als Begründungsmodell 47 ff.
- Bayes'sche Fassung des – 40
- deontologischer – 63 ff., 110
- eigenorientierter – 46 ff., 117
- ethischer – 53 ff., 76, 132, 138, 146, 149, 186
- Gerechtigkeits- 66
- kollektive Selbstaufhebung des – 164
- Koordinationsprobleme des 186 ff.
- Maximierungsforderung des – 96
- Probleme des – 130, 145, 148, 165, 168, 183
- prozedurale Komponente 143
- Rationalitätsmodell des – 11, 16, 47 ff., 79 ff., 93 ff., 109 ff., 128 ff., 152 ff., 167 f., 170 ff.
- Selbstaufhebung des – 152

- strikter (s. a. strikter Konsequentialismus) 75 f., 91, 113 ff., 128 ff., 158 ff.
- teleologischer - 110
- Unvereinbarkeit mit normativen Institutionen 144
- Universalisierbarkeit des - 174 f.
konsequentialistische Abschwächung kantischer Moral 182
- Begründung für Handlungen 173 ff.
- Ethik 73 f., 76, 79, 90, 96, 108, 130, 142
- Gesellschaft 93, 137
- Moral, Idealtypus 151
- Optimierung 109
- Theorien 43 ff., 67 f., 84, 86, 87, 90, 111 f., 130, 145, 174
konsequentialistischer Akteur, rationaler 137, 146, 164
konsequentialistisches Entscheidungskriterium 155, 168
- Paradigma 46, 67, 76
Konsequenzen 12, 17, 19, 50 ff., 76, 77, 138
konsequentialistisch rationale Akteure 42 ff., 153 f., 174 f.
konsequentialistische Rationalitätskonzeption 36 ff., 152 ff.
Konstruktivismus 143 ff., 174, 188
Konventionen 97, 99, 188
Kooperation 80, 112 ff., 115, 125, 137, 127 f., 145, 153, 155 f., 163, 170, 185 ff.
- Probleme der - 127 ff., 153
- Stabilität gesellschaftlicher - 137
Kooperationsbedingungen 112 f., 115 ff., 125 ff., 142, 145, 153
- starke - 112 f.
- schwache - 113, 142
Kooperative Strategie, Wahl der 155, 183 ff.
kooperatives Verhalten 156, 170
Koordination 95, 113, 127, 129, 152, 165, 170, 184, 188
- interpersonelle 165
- intrapersonelle 165, 170
Kritik zweiter Ordnung 139 ff., 143 ff., 147

Lebensform 186 ff.
- gestaltung 186 ff.
- welt 77
Lerntheorie 84
lexikographische Ordnung 63
Liberales Pradoxon 103 ff.
Liberalitätsbedingungen 102 ff., 107 f.
Logik kollektiver Entscheidungen 69, 103, 130, 132

Mackie-Variante 155 ff.
Manipulation 136, 138
methodologische Theorien 139
Metrisierungs-Theorem 40
Monotonieaxiom 179 f.
moralische Person 30, 62, 77, 79, 82, 96 f., 111, 147, 151, 154, 158 ff., 165, 174 f., 188
Motivationsethik 53 ff., 84

naturalistische Ethik 79 ff.
naturalistischer Fehlschluß 79, 172
nicht-axiologische Pflichtenethik 180
nicht-konstruktivistische Ethik 144
Normalschema, konjunktives 21
normativ-axiologische Theorien 80, 82 f.
normative Begründungen 171 ff.
- Institutionen 144, 151, 177
- Prädikate 81
- Urteile 79, 81, 161
normativ-ethische Theorien 44, 65 ff., 138 f., 151, 154, 161 f., 172
normativische Begründungen 176
Normensystem, deontologisches 132
Nutzenfunktionen 44 f.
Nutzenmaximierung 115
Nutzentheorem 39 f., 44

objektive Begründungen 172 ff.
- Kriterien 81 ff.
- Pflichten 96
öffentliches Wohl 76
Ökonomie, klassische und neoklassische 46
One-shot-PD-game 183
Optimierung 128, 165 f., 182
Optimierungsmodell 182
Ordnungen, strikt individuelle 135
Ordnungseigenschaften, von Präferenzen 69

Pareto-Kriterium 72 f., 101 f., 104, 106, 135
performative Äußerung 28
Permutationen 69
Person, allwissende 126
- Freiheit der - 170 ff.
- ideal moralische - 62, 76, 83, 91, 111 ff., 141 ff., 154, 174 f.

- ideal rationale – 36 ff.
- Identität der – 58
- Integrität der – 57, 92 ff., 151
- Interesse der – 74, 76, 171
- moralische – 30, 77, 79, 82, 96 f., 111, 147, 151, 158 ff., 165, 188
- rationale – 39, 43, 47, 154, 178 ff.
- reale – 37, 45
- reduktionistische Theorie der – 58 ff.
- willensschwache – 170
- Zustände der – 17, 19, 46

Pfadunabhängigkeit 136
Pflichtenethik 83
Positionen 69 f.
Präferenzen 18, 37 ff., 50 f., 69 ff., 101, 104 ff., 107 ff., 132 ff., 178, 185 ff.
- bedingte – 18
- Bedingungen von – 37 ff.
- Charakterisierung der – 38 ff.
- eigenorientierte – 71
- erweiterte – 71 f.
- individuelle – 69 ff., 105
- paternalistische – 109
- persönliche – 50 f.
- Reflexivität von – 69
- strategische Verfälschung der – 136
- subjektive – 101, 107, 178
- Transitivität von – 69
- Vollständigkeit der – 69
- Zuordnung von – 71

Präferenzorientierung 104, 106
Präferenzrelationen 43, 69, 71 f., 104 ff., 132 ff., 178
- individuelle 132
- kollektive 135
- subjektive 107 f., 178

Präferenzschwellentheorie 124, 125
Präferenzstrukturen 69, 132, 135, 138, 185 ff.
pragmatischer Imperativ 180
praktische Vernunft 158 ff.
Prärogative 74 ff.
primäre Handlungsmotivation 91 f.
- Regel 95 ff.
- Ziele 148

probabilistische Auffassungen 83
- Einstellungen 26

propositionale Einstellungen 43 ff.
propositionaler Gehalt 25, 26
psychologische Ethik 82
Publizitätskriterium 163
punktuelle Optimierung 132, 165
punktuelle Rationalität 129
Ramsey-Kohärenz 182 ff.
rationale Person 39 f., 43 ff.
Rationalität 11 ff., 18, 37, 47 ff., 53 ff., 79 ff., 93, 109 ff., 118 ff., 151 ff., 161 ff., 185 ff.
- Entscheidungs- 37
- ideale Theorie der – 156
- Kriterium der – 18
- punktuelle – 129
- strukturelle – 128, 169

Rationalitätskonzeptionen 47 ff., 130, 154, 157, 164 f., 185 ff.
- des eigenorientierten Konsequentialismus 47 ff.

Rationalitätskriterien 47 ff., 161 ff.
- modell 148, 151, 163

Rechte, individuelle 60, 95 f., 186 f.
rechtliche Normen 165 f.
Rechtspositivismus 99
Reduktionismus 58 ff.
Reduktionsbedingung 180
Reflexivität 37 f.
Regel kollektiver Entscheidungsfindung 101, 147
Regelkonformität 146
Regeln 50 ff.
Regelutilitarismus 83, 85, 116 f.
relativistische Theorien 115
revealed preference Konzept 184
rigide deontologische Theorien 92
rigide unpersönliche ethische Theorie 90 f.

Sachverhalte 21 ff.
- generische 21 f.
- singuläre 21
- Repräsentanz von -n, durch Sprechakte 25

Sanktionen 145, 188
schwache Kooperativität 113, 142
sekundäre Regeln 95 ff., 99, 105, 137
Selbstbindung 187 f.
Selbstkontrolle 166 ff., 187 f.
Shaftesbury-Variante 155 ff.
Sittlichkeit 92 f., 99, 130, 145
soziologische Ethik 79, 82
Spieltheorie 11, 36, 163
Sprechakttheorie 25 ff.
Stetigkeitsaxiom 179 f.
Strategien 123, 128, 136 ff., 151, 164, 183 ff.
strikter ethischer Konsequentialismus 54, 75 ff., 91 ff., 113 ff., 128 ff., 158 ff.

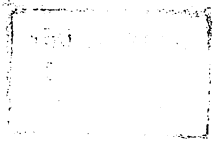
strikter Konsequentialismus 66ff., 113, 115, 117, 120ff.
- deontologische Interpretation 115f.
- kollektive Maximierung bei Interaktion 115ff.
Konsistenz des strikten - 122ff.
- Koordinationsprobleme im - 113, 120ff.
- Unvereinbarkeit mit Integrität der Person 92ff.
- Vereinbarkeit von individueller und kollektiver Maximierung 122ff.
- Vereinbarkeit von Kooperativität 115ff.
strukturelle Irrationalität 166ff.
- Optimierung 165, 169ff.
- Rationalität 128, 169ff.
Strukturkonformität 128
strukturensprengende Wirkung punktueller Optimierung 166
subjekt-zentrierte Rationalitätskriterien 158, 161, 162
subjektive Begründung 172ff.
- Handlungsgründe 174
- Handlungsziele 172
- Moral 82, 96
- Präferenzen 107, 178
- Wahrscheinlichkeitsfunktion 178
- Wertfunktion 80ff.
- Wertorientierung 161
subjektivistisch konsequentialistische Theorie 67ff.
subjektivistische Ethik 66, 81ff., 107, 163ff.
- Werttheorie 66
Substituierbarkeitsbedingung 180
Substitutionsaxiom 179
supererogative Handlungen 182
Sympathie, erweiterte 67ff.
synchronisches Vorgehen 128

Tabu-Ethik 144
technischer Imperativ 180
teleologische Ethik 77, 170
teleologische Theorien 63ff.
teleologischer Konsequentialismus 110
Theorem der Strategieanfälligkeit 135
Theorie der Handlungsbegründung 187
- der utilitaristischen Verallgemeinerung 85
- des außermoralisch Wertvollen 85
Transitivität 106
Theorie der kollektiven Entscheidungen 68ff.
Theorie des richtigen Handelns 64ff.
Transformation, lineare 40, 178
Transitivität, als Bedingung von Präferenzen 38
Tugend- und Motivationsethik 64, 84, 86f., 111

Überbordende Theorie der Verantwortlichkeit 90ff.
Universalisierbarkeit 174ff.
universelle Handlungsmotivation 91
Unterschiedsprinzip 62
Utilitarismus 54ff., 62, 74f., 80, 85, 91, 113f., 117, 139, 144, 146, 174
- als Paradigma konsequentialistischer Ethik 54ff.
- anthropozentrische Variante des - 56
- Entkräftung der -Einwände des - 117
- Handlungs- 54, 57, 113
- hedonistischer Handlungs- 74
- Idealtypus des - 54ff.
- klassischer - 74, 75
- Regel- 54, 83, 85, 116f.
- Varianten des - 117

Verdienstkriterien 63
Verhalten 48, 79, 85, 92, 152, 156ff.
- konsequentialistisch rationales - 152
- kooperatives - 156
- strikt konsequentialistisch motiviertes - 92
Vernunft, praktische 158ff.
Verpflichtung, moralische 86, 146
Verpflichtungsgrundsatz 64
- urteile 84ff., 111
Verteilungsgerechtigkeit 62ff., 117
Vertragsargument, kantische Fassung 96

Wahrhaftigkeitsregel 50f.
Wahrscheinlichkeiten 18, 20, 36ff., 178
Wahrscheinlichkeitsfunktion 178
- verteilung, kohärente 36ff.
- verteilung, subjektive 36ff.
Wahrscheinlichkeitsverteilungen 38ff.
Welten, mögliche 52
Weltverläufe 20, 22f., 36, 42, 167, 177
Weltzustände 17f., 21, 23, 42, 51f., 117, 179, 182
- relative Ungerechtigkeit der - 117
- vollständige Beschreibung von -n 51f.

Wertaggregationen 122ff.
Wertannahmen 117
Wertfunktion 19, 49, 67, 77, 90ff., 100, 116f., 127, 129, 138, 148, 154f., 166ff., 170, 177
- individuelle Prägung der - 127
- individuelle - 138
- interpersonell invariante - 127, 155
- interpersonell invariante - 91
- kollektiv irrationale - 166
- Maximierung der - 116
- Modifikation der - 77
- nicht Maximierung der - 117
- optimale - 170
- quantitative - 67
- präferenzorientierte - 105
- strikt-konsequentialistische - 127
- subjektive - 19, 54, 100
- unpersönliche - 92
- Zeitintegral einer - 168
Werttheorie 59f., 66
Wertorientierung 161
Wissen 82ff.
- Begrenztheit des 84
- vollständiges 82
Wohlergehen 74
Wohlfahrtsfunktion 107, 133
wohlgeordnete Gesellschaft 187
Wünschbarkeitsfunktionen 44
Wünsche 24, 42, 45, 160

Zeitlogik 21
Ziele, primäre 148f.
Zustand, anonym gesellschaftlicher 68f.
- gesellschaftlicher - 68f.
- intentionaler - 34
Zustände 20ff., 36, 50, 62, 67, 85
- assertive - 26
- direktive - 26
- gesellschaftliche - 62ff.
- intentionale - 25ff., 85
- mentale - 26, 29, 85